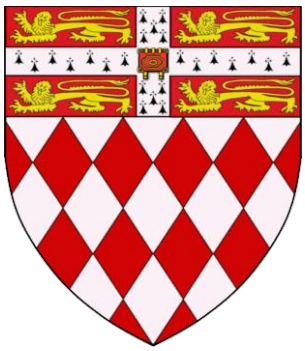
Structural and Biophysical Characterisation of Denatured States and Reversible Unfolding of Sensory Rhodopsin II



Yi Lei Tan

Fitzwilliam College

Department of Biochemistry

University of Cambridge

September 2018

A dissertation submitted for the degree of Doctor of Philosophy

Preface

This dissertation is a summary of the research carried out in the Department of Biochemistry, University of Cambridge, UK, between October 2014 and September 2018. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution. This dissertation does not exceed the prescribed word limit for the Degree Committee of the Faculty of Biology.

Part of the work in Chapter 2 and Appendix A has been published in the following article:
Tan YL, Mitchell J, Klein-Seetharaman J, Nietlispach D (2018) Characterization of Denatured States and Reversible Unfolding of Sensory Rhodopsin II. *J Mol Biol* pii: S0022-2836(18)30312-7. doi: 10.1016/j.jmb.2018.07.031

Abstract

Our understanding of the folding of membrane proteins lags behind that of soluble proteins due to the challenges posed by the exposure of hydrophobic regions during *in vitro* chemical denaturation and refolding experiments. While different folding models are accepted for soluble proteins, only the two-stage model and the long-range interactions model have been proposed so far for helical membrane proteins.

To address our knowledge gap on how different membrane proteins traverse their folding landscapes, Chapter 2 investigates the structural features of SDS-denatured states and the kinetics for reversible unfolding of sensory rhodopsin II (pSRII), a retinal-binding photophobic receptor from *Natronomonas pharaonis*. pSRII is difficult to denature, and only SDS can dislodge the retinal chromophore without rapid aggregation. Even in 30% SDS (0.998 X_{SDS}), pSRII retains the equivalent of six out of seven transmembrane helices, while the retinal binding pocket is disrupted, with transmembrane residues becoming more solvent-exposed. Folding of pSRII from an SDS-denatured state harbouring a covalently-bound retinal chromophore shows deviations from an apparent two-state behaviour. SDS denaturation to form the sensory opsin apo-protein is reversible. This chapter establishes pSRII as a new model protein which is suitable for membrane protein folding studies and has a unique folding mechanism that differs from those of bacteriorhodopsin and bovine rhodopsin.

In Chapter 3, SDS-denatured pSRII, acid-denatured pSRII and sensory opsin obtained by hydroxylamine-mediated bleaching of pSRII were characterised by solution-state NMR. 1D ^{1}H and ^{19}F NMR were first used to characterise global changes in backbone amide protons and tryptophan side-chains. Residue-specific changes in backbone amide chemical shifts and peak intensities in 2D [^{1}H,^{15}N]-correlation spectra were analysed. While only small changes in the chemical environment of backbone amides were detected, changes in backbone amide dynamics were identified as an important feature of SDS- and acid-denatured pSRII and sensory opsin. ^{15}N relaxation experiments were performed to study the backbone amide dynamics of SDS-denatured pSRII, reflecting motions on different timescales, including fast fluctuations of NH bond vectors on the ps–ns timescale and the lack of exchange

contributions on the μs timescale. These studies shed insight on differences in the unfolding pathways under different denaturing conditions and the crucial role of the retinal chromophore in governing the structural integrity and dynamics of the pSRII helical bundle.

Hydrogen bonds play fundamental roles in stabilising protein secondary and tertiary structure, and regulating protein function. Successful detection of hydrogen bonds in denatured states and during protein folding would contribute towards our understanding on the unfolding and folding pathways of the protein. Previous studies have demonstrated residue-specific detection of stable and transient hydrogen bonds in small globular proteins by measuring $^{1}J_{NH}$ scalar coupling constants using NMR. In Chapter 4, different methods for measuring $^{1}J_{NH}$ scalar coupling were explored using RalA, a small GTPase with a mixed alpha/beta fold, as proof-of-concept. Detection of hydrogen bonds was then attempted with OmpX, a beta-barrel membrane protein, both in its folded state in DPC micelles and in the urea-denatured state. While $^{1}J_{NH}$ measurement holds promise for studying hydrogen bond formation, further optimisation of NMR experiments and utilisation of perdeuterated samples are required to improve the precision of such measurements in large detergent-membrane protein complexes.

Naturally-occurring split inteins can mediate spontaneous *trans*-splicing both *in vivo* and *in vitro*. Previous studies have demonstrated successful assembly of proteorhodopsin from two separate fragments consisting of helices A–B and helices C–G via a splicing site in the BC loop. To complement the *in vitro* unfolding/folding studies, pSRII assembly *in vivo* was attempted by introducing a splicing site in the loop region of the beta-hairpin constituting the BC loop of pSRII. The expression conditions for the N- and C-terminal pSRII-intein segments were optimised, and the two segments co-expressed. However, the native chromophore was not observed. Further optimisation is required for successful *in vivo* *trans*-splicing of pSRII and application of this approach towards understanding the roles of helices and loops in the folding of pSRII.

Acknowledgements

I would like to express my sincere gratitude to everyone who has made my Ph.D. journey and time in Cambridge memorable and enjoyable.

First and foremost, I express deep gratitude to my supervisor, Daniel Nietlispach, for his insightful guidance, invaluable advice, unwavering support and encouragement throughout my Ph.D., and for patiently helping me to understand the theory and practice of biomolecular NMR spectroscopy.

I would particularly like to extend heartfelt thanks to Judith Klein-Seetharaman and James Mitchell (University of Warwick) for an extremely fruitful collaboration, for helping me navigate the incredibly exciting field of membrane protein folding, and especially to James for generously writing the *R* codes used for the work in Chapter 2.

Special thanks must go to Mark Bostock for introducing me to pSRII, kind assistance with NMR experiments, Linux and coding, and helpful comments on parts of this thesis. Special thanks to Sandra Berndt for her guidance on the purification and refolding of OmpX. I would like to thank Barbara Hogan, who has worked very hard to kick off the intein project during her summer project in the Nietlispach lab.

I also thank all other past and current members of the Nietlispach lab – Yvonne Yu, Rowina Westermeier, Duncan Crick, Prashant Kumar, Andras Solt, Henry Chien, Niclas Frei and Aditi Tandale – for their kind support and friendship.

I am grateful to my advisor, Helen Mott, my first year viva examiner, Bill Broadhurst, and my Graduate Thesis Panel, Simone Weyand, Marko Hyvonen and Ben Luisi, for their constructive advice on my Ph.D. project.

Thanks to Katherine Stott and Joe Maman for their kind guidance on using various equipment in the Biophysics facility. Thanks to Laura Itzhaki for helpful discussions on protein folding. Thanks to Wayne Boucher for his help in resolving issues with using CCPN Analysis and Azara. Thanks also to Jules Griffin and Srinjan Basu for helpful discussions on principal component analysis.

Throughout my time in Cambridge, Fitzwilliam College has been a wonderful environment in which to live and study. This work was funded partly by the Cambridge Commonwealth, European & International Trust, along with generous provision of

conference stipend by the Cambridge Trust, Cambridge Philosophical Society and Fitzwilliam College.

I am very grateful to friends from Cambridge and back in Hong Kong for much encouragement and support. Special thanks to George Sophocleous, Hilary Wong, Daisy Luff and Joyce Ratti for many fun times and laughter shared over meals, and to Bob Maison and Yasmin Surani for the delightful squash sessions and unforgettable jokes.

I would like to thank Marek Tyl, Piotr Gierszewski, Helen Fox, Karolina Zapadka, Yoshiro Shiba and Alexander Schubert for inspiring and providing me with valuable opportunities to gain preliminary insights into the exciting realms of technology innovation, entrepreneurship and partnerships through the Innovation Forum. Many thanks also to all past and current members of the Innovation Forum Cambridge team, especially to Arfa Karani, Mo Zhao, Hakan Bagci and Shuang Sun, for numerous stimulating projects and discussions.

Finally, I express my most sincere gratitude and appreciation to my family for a lifetime of unconditional love, support and encouragement.

Table of Contents

List of Figures

List of Tables

Abbreviations

7TM	Seven transmembrane
λ_{max}	Maximum absorption/emission wavelength
τ_c	Rotational correlation time
X_{SDS}	Molar fraction of SDS
AIC	Akaike Information Criterion
AP	Antiphase
BMRB	Biological Magnetic Resonance Bank
c7-DHPC	Diheptanoylphosphatidylcholine; 1,2-diheptanoyl-sn-glycero-3-phosphocholine
CCR	Cross-correlated relaxation
CD	Circular dichroism
CPMG	Carr-Purcell-Meiboom-Gill
CSA	Chemical shift anisotropy
CTAB	Cetyltrimethylammonium bromide
DD	Dipole/dipole interaction; dipolar interaction
DDM	n-Dodecyl β-D-maltoside
DFS	Dynamic frequency shift
DPC	Dodecylphosphocholine
E.COSY	Exclusive correlated spectroscopy
FID	Free induction decay
FRET	Förster resonance energy transfer
GdmCl	Guanidinium hydrochloride
H-bond	Hydrogen bond
H/D	Hydrogen/deuterium
HSQC	Heteronuclear single quantum correlation
INEPT	Insensitive nuclei enhanced by polarisation transfer
IP	In-phase
Int^C	C-terminal intein fragment
Int^N	N-terminal intein fragment
J-coupling	Scalar coupling
k_f	Rate constant for protein folding
k_u	Rate constant for protein unfolding

LDAO	Lauryldimethylamine-N-oxide
MRE	Mean residue ellipticity
NDSB	Non-detergent sulfobetaine
NDSB-201	3-(1-Pyridinio)-1-propanesulfonate
NDSB-256	3-(Benzyldimethylammonio)propanesulfonate
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
Npu DnaE	*Nostoc punctiforme* DnaE
Omp	Outer membrane protein
PEG	Polyethylene glycol
PIPES	Piperazine-N,N′-bis(2-ethanesulfonic acid)
PSB	Protonated Schiff base
pSRII	*Natronomonas pharaonis* sensory rhodopsin II
R_1	Longitudinal relaxation rate constant
R_2	Transverse relaxation rate constant
RalA	Ras-like (Ral) protein A
RF	Radiofrequency
S^2	Order parameter
SDS	Sodium dodecyl sulphate
smFRET	Single-molecule FRET
SMFS	Single-molecule force spectroscopy
SNR	Signal-to-noise ratio
SO_{390}	Denatured apo-sensory opsin
SOFAST	Band-selective optimised flip-angle short transient
SR_{440}	Unfolded pSRII with all-*trans* retinal chromophore remaining covalently-bound
TFA	Trifluoroacetic acid
TFE	2,2,2-trifluoroethanol
Triton X-100	Polyoxyethylene octyl phenyl ether
TROSY	Transverse relaxation optimised spectroscopy
TSP	Trimethylsilylpropanoic acid
Tween-20	Polyoxyethylene (20) sorbitan monolaurate
UV/vis	UV/visible

1 Introduction

1.1 Protein folding: lessons from soluble proteins

Protein folding is a process of molecular self-assembly during which a polypeptide chain undergoes numerous conformational changes to form its functional structure. Understanding the mechanisms of protein folding has profound implications on biological research and medical applications. Misfolding, or failure of a protein to fold into its functional conformation, can lead to disruption of regulated cell functions and diseases through various mechanisms, including loss of function of the misfolded protein, cellular stress responses and cell death triggered by toxic accumulation of misfolded protein [1,2].

Studying how a protein folds to its native structure is far more complex than solving classical chemical mechanisms. While changes in only a small number of covalent bonds are involved in a chemical reaction, thousands of weak non-covalent interactions are being made or broken simultaneously during protein folding. Moreover, protein folding intermediates are very difficult to trap because of their unstable nature. Research into the protein folding problem has been ongoing for more than 50 years, during which different folding mechanisms of small, water-soluble proteins have been proposed and become well-characterised. By contrast, progress in understanding membrane protein folding were slower due to various experimental challenges [3], including the necessity of a membrane mimetic environment for *in vitro* studies and difficulties in finding suitable conditions for reversible unfolding.

It has been known since 1961 that proteins fold spontaneously, and that the amino acid sequence encodes the final folded structures and information on how the protein folds [4]. Anfinsen formed the thermodynamic hypothesis of protein folding. It was postulated that the native conformations of proteins are at global free energy minima relative to all other states. Under thermodynamic control, folding is independent of the starting conformation and the pathway. Experimental evidence cited in support of the thermodynamic hypothesis is that the folding/unfolding reactions of many small proteins are reversible [5,6].

The proposal of the "Levinthal's paradox" in the late 1960s [7] led to the conclusion that proteins must fold via defined pathways (i.e. kinetic control), as an unbiased search through an astronomical number of conformations, which are theoretically available to the unfolded state of a protein, would take too long. Under kinetic control, the protein traverses a convoluted energy surface with multiple energy minima. Whether the protein reaches a global or local energetic minimum depends on the height of the energy barriers, the folding pathway and the starting conformation.

The debate between these two viewpoints has continued, with numerous experimentalists and theoreticians investigating whether proteins reach their global energy minimum in a pathway-independent manner under thermodynamic control, or follow a specific pathway to either a local or the global minimum under kinetic control. Levinthal's work initiated the search for folding pathways and characterisation of intermediate states in order to understand how an unfolded protein finds its native structure amongst the large number of possible conformations.

1.1.1 Thermodynamics of protein folding

The overall protein folding process must assume a negative change in Gibbs free energy in order to be spontaneous. The change in Gibbs free energy is a function of change in enthalpy (ΔH) and change in entropy (ΔS) at a given temperature (T):

$$\Delta G = \Delta H - T\Delta S \tag{1.1}$$

The denatured state has considerable conformational freedom, and thus high entropy (S) based on the Boltzmann equation:

$$S = k \ln W \tag{1.2}$$

where W is the number of accessible states and k is the Boltzmann constant.

The denatured states of a protein constitute a collection of conformations with very similar energies and are in rapid equilibrium with one another. Many hydrophobic side chains are exposed to solvent, and hydrogen bond donors and acceptors are hydrogen-bonded to surrounding water molecules. The entropy and enthalpy of water molecules are lowered as

individual molecules have less freedom and hydrogen bonds are made with the denatured protein.

Under denaturing conditions, the denatured states are more thermodynamically favoured than other states. Molecular dynamics simulations suggest that aqueous denaturants, such as guanidinium hydrochloride (GdmCl) and urea, stabilise the denatured states by facilitating the solvation of their nonpolar surfaces [8]. Single-molecule fluorescence study shows that the GdmCl-denatured states of IgG-binding domain of protein L expand with increasing GdmCl concentrations [9]. Calculation of folding free energies from single-molecule data shows linear change in free energy across all denaturant concentrations, suggesting a gradual thermodynamic stabilisation of the denatured states in the presence of higher denaturant concentrations. Conformational entropy has a large contribution to the free energy of folding, as expansion of the denatured states necessitates a large change in entropy [8,9]. Various models have been proposed for the denaturing mechanism of surfactants, as protein-surfactant interactions are too diverse to be accommodated in one model [10]:

(1) a "rod-like" prolate ellipsoidal protein-surfactant aggregate, in which the length depends on the protein molecular weight and the width of ~18 Å corresponds to the alkyl chain length of SDS [11];
(2) the decorated micelle model: the extended denatured protein wraps around a flexible capped cylindrical micelle, and the complex is stabilised by hydrogen bonds between the protein and detergent headgroups [12];
(3) the necklace-and-beads model: the polypeptide chain is considered to be mostly flexible, and micelles could form on different parts of the extended protein. Electrostatic repulsion between individual micelles is a driving force for protein denaturation [13,14].

The same protein may engage in different types of protein-micelle interactions during different stages of protein denaturation (Figure 1.1). Factors which determine the stability and properties of protein-micelle complexes include the protein structure, the type of surfactant, the protein-to-surfactant ratio, and solvent conditions [10].

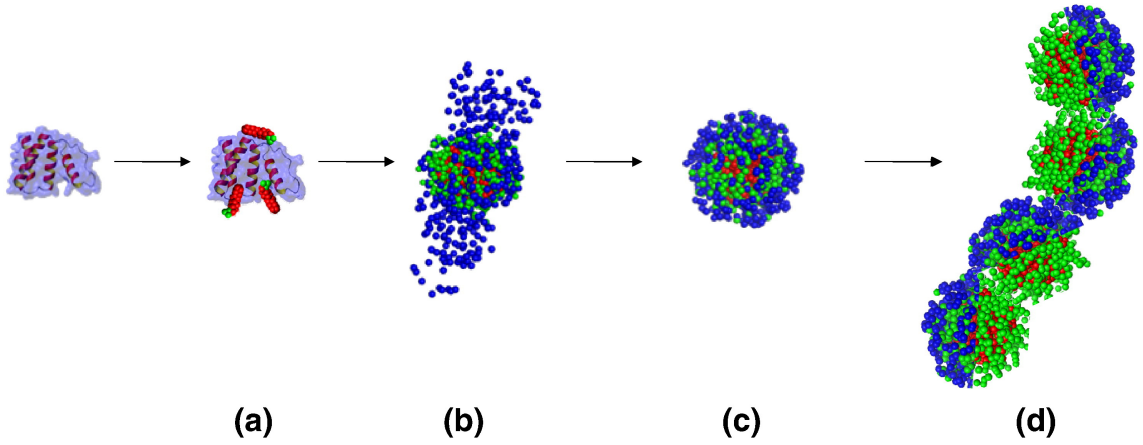


Figure 1.1: Schematic representations of different stages in the denaturation of bovine acyl-coenzyme-A-binding protein (ACBP)
In stage (a), ACBP binds SDS monomers without losing the native structure. Stage (b) involves the formation of a decorated micelle with 37 SDS molecules binding 2 ACBP molecules. In stage (c), further binding of SDS leads to a total of 40 SDS molecules forming a shell-like structure and binding to monomeric ACBP. Stage (d) shows a speculative structure representing the "necklace-and-beads" model which has been proposed for protein interactions with SDS micelles above the CMC. The hydrocarbon tails of SDS are represented with red dots, the headgroup and counterions of SDS are represented with green dots, and protein distribution is represented with blue dots. Figure retrieved from [15].

Protein folding is guided by different driving forces, mainly the formation of hydrophobic interactions and intramolecular hydrogen bonds, and is opposed by the loss of conformational entropy. The folding process also depends on the surrounding environment, including the type of solvent (aqueous environment vs. lipid bilayer), salt concentration, pH, temperature, and the presence of cofactors and molecular chaperones. Hydrophobicity is the dominant driving force in protein folding, leading to compact conformations with nonpolar cores [5] and contributing 1.1 ± 0.5 kcal/mol for every CH_2 group that is buried upon folding [16]. A hydrogen bond contributes 1.1 ± 0.8 kcal/mol to protein stability [16]. Other forces are weaker but can also affect stability. These include disulphide bonds, electrostatic interactions, hydrogen bonds to buried charged side-chains, salt bridges, $n \rightarrow \pi^*$ interactions (electron delocalisation) between carbonyl groups, and gain in conformational entropy of water molecules [16].

The protein loses entropy as it folds, while the enthalpy of the packing of side-chains is favourable and compensates for the low entropy in the native state. The loss of conformational entropy of the protein is a major barrier to protein folding, and is estimated to

be ~2.4 kcal/mol per residue [16]. The marginal stabilities of proteins arise from the small difference between the large driving and opposing forces of protein folding [16]. Water molecules are also released during protein folding. The gain in entropy of water molecules further compensates for the loss of conformational entropy in the protein.

The thermodynamic description of chemical denaturation has traditionally been based on Tanford's transfer model [6,17], which relates the free energy of folding to changes in solvation of amino acid residues when transferred from denaturant to water:

$$\Delta G_{\mathrm{D-N}} = \Delta G_{\mathrm{D-N}}^{\mathrm{H_2O}} - m_{\mathrm{D-N}}[\mathrm{denaturant}] \tag{1.3}$$

where $\Delta G_{\mathrm{D-N}}^{\mathrm{H_2O}}$ is the free energy of denaturation in water, and N and D refer to the native states and denatured states, respectively.

The transfer model assumes that the folding process is represented by only the folded and the denatured conformations (i.e. two-state kinetics).

The slope, $m_{\mathrm{D-N}}$, can be expressed in terms of the free energy of transfer of group *i* from water to a denaturant solution with concentration of 1 M (δg_i), and the fractional change in the solvent-accessible surface area (SASA) of group *i* between the folded and the unfolded states (δa_i):

$$m_{\mathrm{D-N}} = \sum_i \delta a_i \delta g_i \tag{1.4}$$

It is important to note that the free energy of a particular molecule at equilibrium depends only on the state of the molecule and is independent of the pathway by which it is formed. Measurements of the equilibrium distribution of different folding species (native, intermediates, denatured) yield only the relative thermodynamic properties of these species and not the pathway between them [6].

1.1.2 Kinetics of protein folding

The classical models for protein folding are based on simple kinetic models proposed from raw data of single- or multi-exponential time decays of optical properties of the protein after a jump to folding or unfolding conditions [6]. While high-resolution structural information of

the folded state can be obtained by X-ray crystallography or NMR, kinetic folding intermediates have short lifetimes and cannot be isolated for structural studies. Spectroscopic methods are typically used for studying kinetic folding intermediates. These techniques yield valuable kinetics information, but provide little structural information and only probe the average behaviour of the entire protein population.

Two principal factors are required for kinetics studies [18]:

(1) simultaneous population of the native (folded) and denatured states under conditions that can be related to conditions of the native state, for which the free energy of folding is desired; and
(2) conditions in which the reaction is fast enough to reach equilibrium.

Chemical denaturation *in vitro* offers an avenue for accessing selected conformations within the ensemble of structural states adopted by a protein whilst it is folding [19].

Although protein folding is very complex and involves the formation and breaking of many weak non-covalent bonds, this process follows simple rate laws that are governed by a few basic principles.

"Two-state" kinetics refers to a single exponential decay being observed in both the folding and unfolding directions, meaning the data can be fitted by assuming only the presence of a native state (N) and a denatured state (D). The transition state is the ensemble of states which differ slightly from one another in energy around the saddle point of the energy surface. On average, the transition states ensemble is intermediate in structure between the denatured and native states. Denaturants lower the activation energy for unfolding and raise the activation energy for folding. The transition state is stabilised by denaturant with respect to the native structure, but is destabilised relative to the denatured state. The stabilisation energy terms are linear with denaturant concentration ([denaturant]) (Figure 1.2a):

$$\Delta G_{\ddagger-\mathrm{N}} = \Delta G_{\ddagger-\mathrm{N}}^{\mathrm{H_2O}} - m_{\ddagger-\mathrm{N}}[\mathrm{denaturant}] \quad (1.5)$$

$$\Delta G_{\ddagger-\mathrm{D}} = \Delta G_{\ddagger-\mathrm{D}}^{\mathrm{H_2O}} + m_{\ddagger-\mathrm{D}}[\mathrm{denaturant}] \quad (1.6)$$

Many small proteins demonstrate two-state behaviour. A plot of the natural logarithm of the rate constant for unfolding (k_u) against [denaturant] is linear (Figure 1.2b):

$$\ln k_u = \ln k_u^{H_2O} + m_u[\text{denaturant}] \quad (1.7)$$

where $k_u^{H_2O}$ is the rate constant for unfolding in the absence of denaturant obtained by extrapolation and m_u is the constant of proportionality.

The natural logarithm of the rate constant for folding (k_f) is also linear against [denaturant]:

$$\ln k_f = \ln k_f^{H_2O} + m_f[\text{denaturant}] \quad (1.8)$$

The two curves combine to form a V-shaped kinetics curve, also known as the chevron plot (Figure 1.2b):

$$\ln k_{obs} = \ln\left(k_f^{H_2O}\exp(-m_f[\text{denaturant}]) + k_u^{H_2O}\exp(m_u[\text{denaturant}])\right) \quad (1.9)$$

The total free energy of unfolding and the corresponding *m* value can be derived from kinetics data:

$$\Delta G_{D-N}^{H_2O} = -RT\ln\left(k_u^{H_2O}/k_f^{H_2O}\right) \quad (1.10)$$

$$m_{D-N}^{H_2O} = -RT(m_u + m_f) \quad (1.11)$$

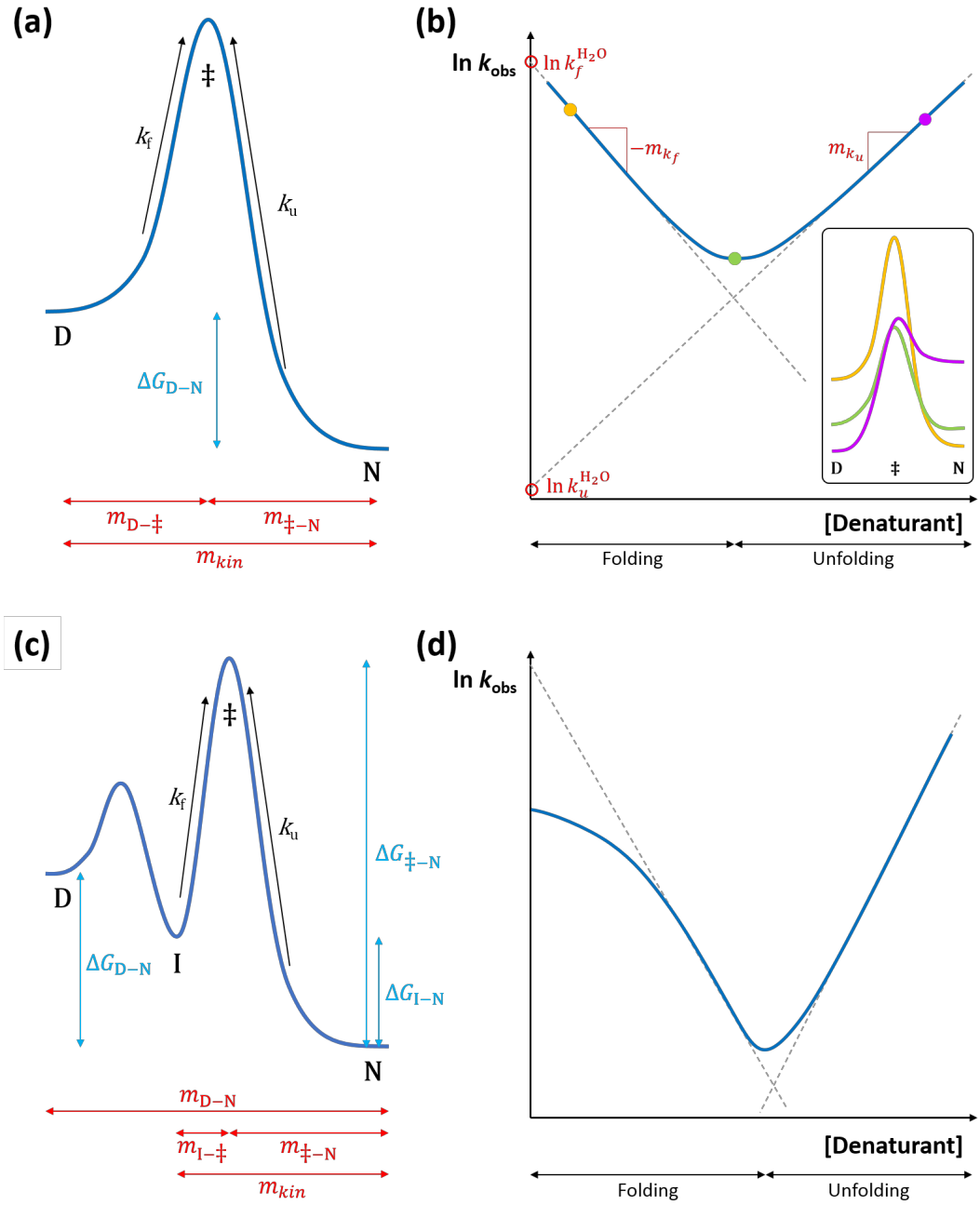


Figure 1.2: Energy diagram and chevron plot for two-state kinetics and three-state kinetics
(a–b) Two state kinetics. **(a)** Energy diagram, showing that the denatured state (D) and the native state (N) have a free energy difference of ΔG_{D-N} and are separated by one transition state (‡) which determines the folding (k_f) and unfolding (k_u) rate constants. **(b)** Two-state kinetics yield linear folding and unfolding arms on the chevron plot. The inset shows the corresponding energy diagrams at three denaturant concentrations (yellow, green and purple dots), illustrating that denaturant stabilises ‡ and D relative to N. **(c–d)** Three-state kinetics with an on-pathway intermediate (I). **(c)** The major transition state is between I and N, and hence ΔG_{I-N} is observed. **(d)** Non-linearity ("roll-over") is observed in the chevron plot because apparent $k_f = k_{I-N} < k_{D-N}$. See main text for information on $\ln k_f^{H_2O}$, $\ln k_u^{H_2O}$ and m values. Figure adapted from [6].

It is essential to measure the refolding arm of the chevron plot down to zero denaturant in order to conclusively demonstrate that there is no deviation from the kinetics in the presence of low concentrations of denaturant during refolding. Exposure of hydrophobic surfaces in denatured states could lead to aggregation, leading to artifacts in the kinetics that could be mistaken for the presence of folding intermediates. The dependence of refolding kinetics on protein concentration must therefore be assessed carefully.

The energy landscape around transition states can be probed by structure-reactivity relationships. Protein folding studies of mutants effectively move the protein along the reaction coordinate. The Tanford β_{T} value provides an index for the position of the transition state along the reaction coordinate and its compactness; it is a measure of the average degree of exposure in the transition state relative to that of the denatured state from the native state:

$$\beta_{\mathrm{T}} = \frac{m_{\mathrm{f}}}{m_{\mathrm{f}}+m_{\mathrm{u}}} = \frac{\sum_i a_{i\ddagger}\delta g_i}{\sum_i a_i \delta g_i} \tag{1.12}$$

where $a_{i\ddagger}$ and a_i are the fractional degrees of exposure in the transition state and denatured state, respectively, and δg_i is the free energy of transfer from water to a denaturant solution with concentration of 1 M.

Protein engineering approaches have been employed to study the structures of transition states and intermediates. ϕ-value analysis involves making mutations to introduce small changes and remove a specific interaction which stabilises the native structure, and measuring the kinetics of folding/unfolding and changes in equilibrium stability ($\Delta\Delta G_{\mathrm{N-D}}$) in different mutants. Structures of transition states and intermediates may be inferred from changes in kinetics and equilibria upon mutation.

$$\phi = \frac{\Delta\Delta G_{\ddagger-\mathrm{D}}}{\Delta\Delta G_{\mathrm{N-D}}} \tag{1.13}$$

The observation of simple two-state kinetics does not necessarily prove that there are no intermediates on the folding pathway. Multistate denaturation equilibria might exhibit apparent two-state behaviour if the intermediates are present at low concentrations or have high energies and are kinetically undetectable.

More complex models are required if multiple exponentials are observed, indicating the presence of more than two states. Three procedures have been described that are effective for detecting intermediates for both apparent two-state kinetics and multi-state kinetics:

1) The ratio of k_u and k_f for a two-state transition must give the same equilibrium constant (K_{eq}) as that measured directly by solvent denaturation determined using the ratio of denatured state concentration ([D]) and native state concentration ([N]):

$$K_{\mathrm{eq}} = \frac{k_u}{k_f} = \frac{[\mathrm{D}]}{[\mathrm{N}]} \tag{1.14}$$

For a three-state reaction in which there is one intermediate state (I), and the rate-limiting step and the major transition state are between I and N, the ratio of the rate constants for the apparent folding step (k_f) and for unfolding in water ($k_u^{H_2O}$) gives the equilibrium constant between I and N, and not between D and N (Figure 1.2c):

$$k_{\mathrm{f}} = k_{\mathrm{I}\to\mathrm{N}} < k_{\mathrm{D}\to\mathrm{N}} \tag{1.15}$$

2) The sum of the forward and reverse m values must be equal to the m value for equilibrium unfolding for a two-state transition (Figure 1.2a):

$$m_{\ddagger-\mathrm{N}} + m_{\ddagger-\mathrm{D}} = m_{\mathrm{D}-\mathrm{N}} \tag{1.16}$$

where all terms are algebraically positive

3) Deviation from linearity ("roll-over") is observed in either/both arms of the chevron plot, where the calculated rate constant(s) derived from a linear extrapolation is much higher than the observed rate constant(s) (Figure 1.2d).

Intermediate states can be on pathway:

$$\mathrm{D} \underset{k_{\mathrm{ID}}}{\overset{k_{\mathrm{DI}}}{\rightleftharpoons}} \mathrm{I} \underset{k_{\mathrm{NI}}}{\overset{k_{\mathrm{IN}}}{\rightleftharpoons}} \mathrm{N} \tag{1.17}$$

or off pathway:

$$\mathrm{I} \underset{k_{\mathrm{DI}}}{\overset{k_{\mathrm{ID}}}{\rightleftharpoons}} \mathrm{D} \underset{k_{\mathrm{ND}}}{\overset{k_{\mathrm{DN}}}{\rightleftharpoons}} \mathrm{N} \tag{1.18}$$

where k_{XY} denotes the rate constant for the forward or the reverse reaction in the equilibrium between states X and Y.

Hence, the formation of N speeds up with the accumulation of I in the on-pathway model, but slows down in the off-pathway model.

According to the on-pathway model, the folding rate is:

$$k_{\mathrm{f}} = f_{\mathrm{I}} k_{\mathrm{IN}} \tag{1.19}$$

where f_{I} is the fraction of the total protein molecules in the I state.

According to the off-pathway model,

$$k_{\mathrm{f}} = (1 - f_{\mathrm{I}}) k_{\mathrm{DN}} \tag{1.20}$$

Hence, in the on pathway model, roll-over in the chevron plot occurs because [20]:

$k_{\mathrm{f}} \propto k_{\mathrm{DI}} k_{\mathrm{IN}} / k_{\mathrm{ID}}$ when $f_{\mathrm{I}} \ll 1$ (i.e. low [denaturant] and $k_{\mathrm{DI}} < k_{\mathrm{ID}}$);

$k_{\mathrm{f}} \propto k_{IN}$ when $f_{\mathrm{I}} \sim 1$.

On the other hand, roll-over in the chevron plot occurs in the off-pathway model because:

$k_{\mathrm{f}} \propto k_{\mathrm{DN}}$ when $f_{\mathrm{I}} \ll 1$;

$k_{\mathrm{f}} \propto (1 - f_{\mathrm{I}}) k_{\mathrm{DN}}$ when $f_{\mathrm{I}} \sim 1$.

Since protein folding kinetics studies often focus on monitoring the disappearance of D or the formation of N, it is very difficult to determine whether the observed "intermediate" is on or off the pathway. In principle, on pathway intermediates cause a lag in the rate of formation of N, whereas off pathway intermediates lead to a burst of formation of N followed by slower folding, as there is initially rapid partitioning of D to I and N followed by slower folding from I to N via D. However, given that the rate constant for the formation of I is much greater than that for N, detection of the lag or burst phase is difficult and requires exquisite data sensitivity. The best method would be to measure the accumulation of a detectable intermediate. However, direct observation of the intermediate state is very challenging in protein folding studies since folding intermediates usually have short lifetimes relative to experimentally-accessible timescales, and the sum of all spectroscopic signals from D, I and N is usually being collected [6].

Multistate kinetics which do not constitute as the typical three-state system have also been observed for a handful of soluble proteins. For example, 86% of hen egg white lysozyme molecules fold via an intermediate due to a kinetic trap at low GdmCl concentrations. This leads to an accumulation of the intermediate and slowing down of the folding rate. The

remaining 14% of the molecules escape this kinetic trap and fold directly from the denatured state to the native state [21]. Complex kinetics are also observed for dihydrofolate reductase (DHFR). Native DHFR is composed of the NADPH-binding conformation and the non-NADPH-binding conformation, and there are separate folding and unfolding pathways for both conformations [22].

Proline *cis–trans* isomerisation also plays an important role in the kinetics and the rate-determining steps of protein folding [23]. The equilibrium constant for the normal peptide bond in proteins favours the *trans* conformation by a factor of 10^3–10^4 due to interactions of the $C\alpha_i$ and $H\alpha_i$ with $C\alpha_{i+1}$ and $H\alpha_{i+1}$ in the *cis* conformation, along with the more favourable electrostatic interactions between O_i and C'_{i+1}. Proline groups exist essentially in either the *cis* or the *trans* form in native proteins because they are almost isoenergetic. At ambient temperatures, proline isomerisation occurs with half-life of 10–100 s, which is generally much slower than conformational folding [6]. An equilibrium mixture of the *cis* and *trans* forms is also present in unfolded proteins. Several slow refolding phases are observed as the different isomers fold in parallel. Classic hallmarks of proline isomerisation include slow reaction with an activation energy of 20 kcal/mol, possibly due to the partial double-bond character of the peptide bond, and catalysis by peptidylprolyl isomerases [23].

The kinetic and structural simplicity of proline isomerisation (which is essentially conformational change in the peptide bond dihedral angle ω) provides an assay for conformational changes and folding pathways [23]. Ribonuclease (RNase A) has four prolines, of which Pro-93 and Pro-114 are *cis* in the native state, and Pro-42 and Pro-117 are *trans* [24]. GdmCl-unfolded RNase A contains a heterogeneous mixture of *cis*/*trans* proline isomers. Double-jump refolding experiments on the wild-type and proline-to-alanine mutants of RNase A involving a first jump to high GdmCl concentrations followed by a second jump to dilute the GdmCl enabled the contributions of the three essential prolines to the five observed refolding phases to be elucidated [25,26]. It is postulated that non-native proline isomers induce localised disruptions by altering local peptide-peptide hydrogen bonds (local secondary structure) which are essential for subsequent steps of protein folding [23]. For example, the native *cis* isomer of Pro-93 naturally forms a β-turn, which may then act as the folding nucleus, catalysing the formation of the two-stranded β-hairpin [23,27]. Conversely,

the non-native *trans* isomer of Pro-93 disrupts the native β-turn, thus inhibiting the formation of the adjacent β-hairpin motif [23,28]. In this case, formation of the native hairpin would require nucleation by longer range interactions, and might also compete against the formation of non-native local structure, both of which would lead to slow conformational folding [23].

1.1.3 Classical view: Folding pathways

Three classical models have emerged from protein folding research on soluble proteins [6,29] (Figure 1.3):

(1) framework model [30,31],
(2) hydrophobic collapse [32–34], and
(3) nucleation condensation [35,36].

At one extreme, the framework (or hierarchic) model postulates stepwise formation of structure beginning with rapid formation of local secondary structures, which function as a scaffold for subsequent interactions between different secondary structure elements to form more advanced folding intermediate(s) and the tertiary structure. Diffusion-collision involves the formation of "microdomains" of secondary structures, which diffuse, collide and coalesce to form the native tertiary structure [37]. The framework model implies the existence of folding intermediates. Early intermediates have secondary structures fluctuating around their native positions; late intermediates are compact, have native-like architecture, but lack the unique native tertiary structure. Each step forward in the folding process provides stabilisation of the major structural elements formed at the previous stage.

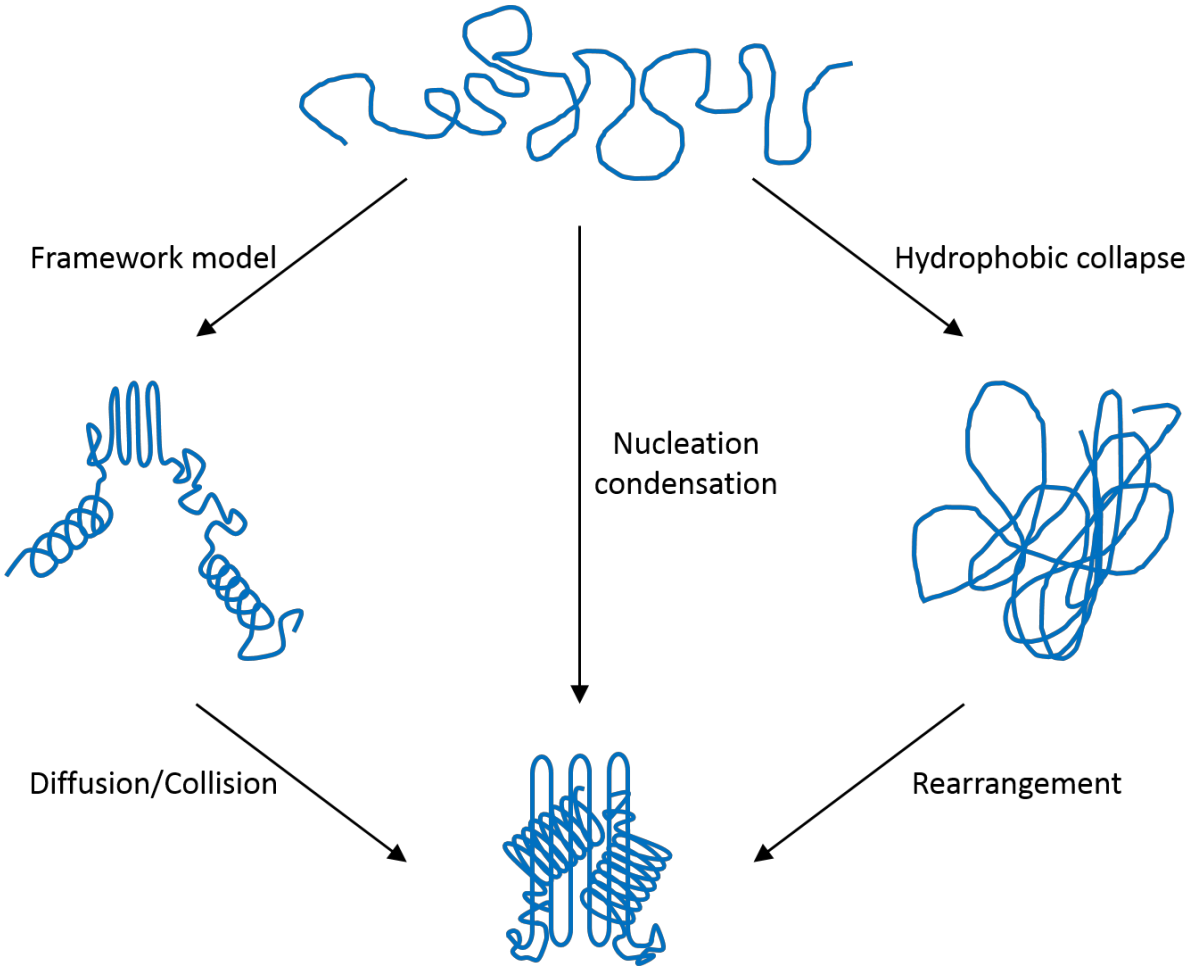


Figure 1.3: The three classical mechanisms for protein folding
In the framework model, early intermediates feature secondary structures fluctuating around their native positions. These elements diffuse and collide to form late intermediates with compact, native-like architecture and well-formed secondary structure elements. The hallmarks of the nucleation condensation model are the lack of folding intermediates and no fully-formed secondary or tertiary structures in the transition state. Hydrophobic collapse involves the formation of a compact molten globule state followed by its rearrangement to form the native structure. Figure adapted from [6].

The framework model is exemplified by *Bacillus amyloliquefaciens* ribonuclease (barnase). Barnase has an α + β structure with a major helix, two smaller helices, and an antiparallel β-sheet with five strands. There is a well-packed hydrophobic core formed by the major helix packing against one face of the sheet. All the peptidyl-proline bonds are in the *trans* conformation [38]. Detection of amide-to-amide NOEs by NMR shows that the denatured ensemble contains helical structure in the C-terminal portion of helix 1 and non-native hydrophobic clustering between β-strands 3 and 4 [39]. Molecular dynamics simulations further suggest the presence of tertiary contacts between residues in helix 1 and the

C-terminal portion of the β-sheet, suggesting important roles of long-range interactions in stabilising the helix [40]. The folding of barnase proceeds via an intermediate and the major transition state is between the intermediate and the native state. ϕ-value analysis shows that many mutants of barnase have ϕ-values of close to 0 or 1 for the intermediate and the transition states, indicating that the interactions which are probed by the mutation are either completely absent or fully formed, respectively. The center portions of the β-sheet and the major α helix are nearly completely folded, but the hydrophobic core formed by docking of the major helix onto the sheet is weak. The rate-determining step involves the consolidation and rearrangement of the hydrophobic core [41].

Interestingly, the folding nucleus need not be an essential initiation site. The major helix of barnase is subjected to comparative studies on the effects of Ala vs. Gly mutations (Ala→Gly scanning) of the residues at solvent-exposed positions. As the helix is destabilised, it becomes less structured in the major transition state and is consolidated only late in folding. Conversely, the structure of the rest of the protein becomes formed more completely. This suggests that the structure of the transition state moves gradually around the potential energy surface, and the energy levels are very close together in order to manifest a coordinated effect on the structure upon mutation [42]. The fragment 23–110, consisting of the β-sheet and the remaining α-helices, is capable of partial folding to achieve a catalytically competent state. The full native-like structure is attained upon addition of fragment 1–22 containing the major helix [43]. This behaviour shows that barnase can fold in separate modules, and is consistent with the postulation that the major helix is not an essential initiation site.

At the other extreme, the hydrophobic collapse mechanism proposes that the initial steps in folding involve sequestration of hydrophobic side-chains from surrounding water molecules. This collapsed intermediate is also referred to as “molten globule”. Molten globules are compact, partially folded states, and are characterised by having few tertiary interactions, some secondary structure and a fluctuating hydrophobic core. The acid-induced molten globule of bovine α-lactalbumin has native-like compactness and secondary structure, but has symmetrical environment of aromatic side-chains and faster hydrogen exchange than in the native state [32]. Studies on apomyoglobin [44,45] and cytochrome *c* [46] show that some clusters of helices and other parts of the protein are substantially protected from hydrogen

exchange in the molten globule state. The most stable helices in the molten globule state are largely the same ones which appear at the early stages of protein folding, for example, helices A, G and H in apomyoglobin. Disruptive mutations in helix A/H and G/H packing sites in apomyoglobin lead to substantial destabilisation of both the native structure and the urea-induced molten globule. In comparison, mutations outside helices A, G and H destabilise the native state but have little influence on the stability of the molten globule. These results suggest the importance of tight packing interactions amongst side-chains in stabilising molten globules [47]. After formation of the molten globule state, the protein rearranges within the restricted conformational space. Studies on cytochrome *c* provide evidence for the importance of tight side-chain interactions during the early stages of protein folding. Leu-94 is located at an interface between N- and C-terminal helices. The effects of different mutations of Leu-94 on folding intermediates were compared. Replacement of Leu-94 by similar residues, such as Ile or Val, has little influence on the population of the compact but relatively unstructured intermediate I_C, but leads to a strong decrease in the population of the subsequent partially folded intermediate I_{NC}. On the other hand, mutation to Ala decreases the stability of both intermediates. These results suggest that the early I_C intermediate is stabilised mainly by hydrophobic interactions, while the later I_{NC} intermediate is stabilised by specific interactions mediated by proper packing of the two helices [48].

The nucleation condensation model postulates that neighbouring residues in some parts of the protein sequence would form native secondary structure which acts as a nucleus for stepwise propagation of the native structure. The nucleation site need not be extensively pre-formed in the denatured state and may be in the process of being formed in the transition state. This weak local nucleus is stabilised by long-range interactions to give a large extended nucleus. In the transition state, there must be a critical number of interactions, such that the lowering of enthalpy is more rapid than the loss of entropy for subsequent interactions which are made during folding. This is because the stabilising interactions cluster together and form cooperatively. The choice of residues which constitute the nucleus is somewhat arbitrary, and is conventionally chosen as the residues with the highest ϕ-values.

The nucleation condensation model arose from kinetics studies of chymotrypsin inhibitor 2 (CI2). The native state of CI2 consists of a binding loop, a single α-helix and a mixed parallel

and antiparallel β-sheet. Its four peptidyl-proline bonds are all in the *trans* conformation. The denatured state is relatively expanded and has little residual native structure. CI2 folds and unfolds via simple two-state kinetics, with no intermediates accumulating at equilibrium or in the folding pathway. ϕ-value analysis shows fractional ϕ-values for the majority of sites probed [49]. The interpretation of fractional ϕ-values is given by the Brønsted plot of $\Delta \ln k_f$ against $\Delta \ln K_{D-N}$, where a linear trend is consistent with a narrow ensemble of structures whose average resembles an expanded transition state. This suggests that no region of CI2 is fully formed in the transition state, and interactions in the transition state are weakened compared to those of the native state. The transition state is constructed around an extended, delocalised nucleus consisting of the N-terminal region of the helix and ancillary interactions with residues in the hydrophobic core. Secondary and tertiary structure are formed in parallel as CI2 folds [50].

While there is unlikely to be a single mechanism for protein folding, the three classical folding models point toward a unified scheme, variations of which could be used to describe a large number of pathways [6]. The basic hypothesis is that the folding of small proteins (e.g. CI2) emulates the folding of foldons – basic units which fold quasi-independently in a single cooperative step. Larger proteins can be thought of as comprising multiple foldons so that parallelisation of the search process enables the folding time of the entire molecule to be comparable to that of the smaller individual molecule. The duration of the folding process varies depending on the protein of interest. Under *in vitro* conditions, very small single-domain proteins could be capable of folding in a single step within timescales of milliseconds, or even microseconds, whereas the slowest folding proteins require minutes to hours to fold primarily due to proline isomerisation and formation of obligatory intermediate states. Whether folding proceeds in a concerted (nucleation condensation model) or stepwise (framework model) fashion could depend on the stability of individual foldons within the protein and their ability to fold independently to adopt native-like conformations. It has also been suggested that different folding models might dominate at different stages in folding [51]. Microscopic-level conformational search at the earliest stages of folding enables preliminary folding of individual foldons. Foldon-foldon interactions then collectively provide structural guidance and free energy bias for efficient formation of the native conformation.

1.1.4 Energy landscapes and folding funnels

Energy landscapes, a seminal concept proposed by Frauenfelder, Sligar and Wolynes in 1991 [52], are multidimensional representations of the free energy of different conformations as a function of the degrees of freedom. Each conformation is represented by a point on the landscape. This “New View” for protein folding kinetics emerged as a result of advances in experimental methods and statistical mechanics models (lattice-based representations of chain geometries and interactions, analysed by analytical methods and computer simulations). These experimental and modelling methods recognise that the denatured, transition, intermediate and native states are effectively ensembles of individual chain conformations, and elements of molecular configurations can overlap substantially between different states [53]. Effectively, energy landscapes place more emphasis on conformational ensembles and parallel folding routes and less emphasis on specific structures and pathways. The kinetic process of folding or unfolding a protein is likened to rolling a ball on the energy landscape. Folding is described as a diffusion-like process. The motions of individual chains are asynchronous in the denatured state. Individual chains fold by Brownian type motion – the chains collapse and reconfigure between multiple geometrically similar conformations and follow a general drift from higher to lower energy conformations, ultimately resulting in the same native structure [53].

Different representations of the energy landscape have been used to draw parallels to Levinthal’s paradox and the concept of folding pathways. The Levinthal's paradox, which involves exhaustive sampling of all chain conformations, is analogous to an absolutely flat “golf course” model of the protein potential energy surface (Figure 1.4a). This model is unrealistic, as the protein must lose all of its entropy before any loss in enthalpy. Figure 1.4b shows the “pathway” solution to the random search problem, in which a defined pathway leads from the denatured state to the native conformation. While the concept of a pathway is useful for explaining folding “milestones” such as accumulated intermediates and rate determining steps, it does not enable explanation of minor populations with alternative folding mechanisms or species which are unobservable on experimental timescales.

The “New View” recognises that part of the solution to Levinthal’s paradox is the concept of “folding funnels”, which replaces the pathway concept of sequential events by showing how different folding events can occur in parallel [54]. The funnel is a progressive collection of geometrically similar collapsed structures, of which the native state is more thermodynamically favourable than all other states. The width of the funnel represents the entropy of the system, and the depth represents the enthalpy. Figure 1.4c shows an idealised smooth protein folding funnel. However, the smooth funnel model assumes that kinetic pathways always follow downhill thermodynamic trajectories and does not explain the observation of folding “milestones” such as accumulated intermediates and transition states. This means although the smooth funnel is justified when barriers between states are very low and all conformations are effectively in rapid thermal equilibrium, this model is not justified in scenarios where there are high energy barriers.

Bryngelson *et al*. first explored bumpiness or “frustration” in protein folding landscapes. Rugged folding landscapes are a consequence of the impossibility to simultaneously minimise the energies of all the interactions being formed during the protein folding process [55]. A rough landscape leads to slower folding by limiting the rate at which the protein can find the native state. Depending on the height of the energy barrier, local energetic minima can also act to separate the conformational search into multiple, smaller problems, thereby accelerating the conformational search process.

A detailed representation of the folding funnel was derived from a simulation study of a 60-residue fast-folding helical protein by Onuchic *et al* [56]. The funnel is described by its ruggedness, gradient toward the folded state and configurational entropy. At the folding transition, the energy (enthalpy) loss must equal the temperature multiplied by the entropy loss. Using thermodynamics, the estimated entropy yields the slope of the folding funnel. Measurement of the reconfiguration time of the molten globule enables estimation of the ruggedness of the energy landscape. Folding is driven by an increase in Q, the fraction of correct contacts made in a state i, as the collection of states rapidly interconverts:

$$Q_i = \frac{(\text{number of native pairwise contacts})_{\text{in state } i}}{(\text{total number of pairwise contacts in native state})} \tag{1.21}$$

A cross section of the funnel (Figure 1.4e) illustrates that a fast-folding helical protein has an ensemble of collapsed molten globule states at $Q \sim 0.27$. Nucleation occurs if certain regions experience an increase in Q more rapidly than others. At $Q \sim 0.6$, the incomplete compensation of entropy decrease by enthalpy decrease leads to a sizeable energy barrier (~$3k_BT$). The sizes of energy barriers are described in terms of the product of the Boltzmann constant (k_B) and the absolute temperature (T). According to the Arrhenius equation, the rate constant for a reaction is given as:

$$k = Ae^{-\frac{E_a}{k_BT}} \tag{1.22}$$

where the activation energy per molecule (E_a) is in the same units as k_BT, and A is a constant of proportionality.

The transition state consists of an ensemble of structures, reflecting multiple pathways of protein folding. The entropy continues to decrease until discrete kinetic intermediates appear at $Q \sim 0.7$. The search through these native-like but misfolded structures can be rate-limiting if the thermodynamic bottleneck remains fairly wide and the landscape is still rough at this stage. Hence, folding intermediates are relics of the landscape ruggedness. The native state must be thermodynamically more favourable than the kinetic intermediates in order for the protein to escape these "traps" [54].

The funnel-like shape of the energy diagram, which originates from illustrating the decrease in conformational entropy, often creates a misconception that the funnel "guides" folding to the native state [57]. It must be emphasised again that the decrease in conformational entropy is a major barrier for protein folding and does not aid the polypeptide chain in finding its native state. The major determinant of the folding rate is the free energy landscape. The free energy is the sum of enthalpy and entropy. The free energy barrier to folding manifests as a result of the delicate balance between the favourable decrease in enthalpy and the unfavourable decrease in entropy.

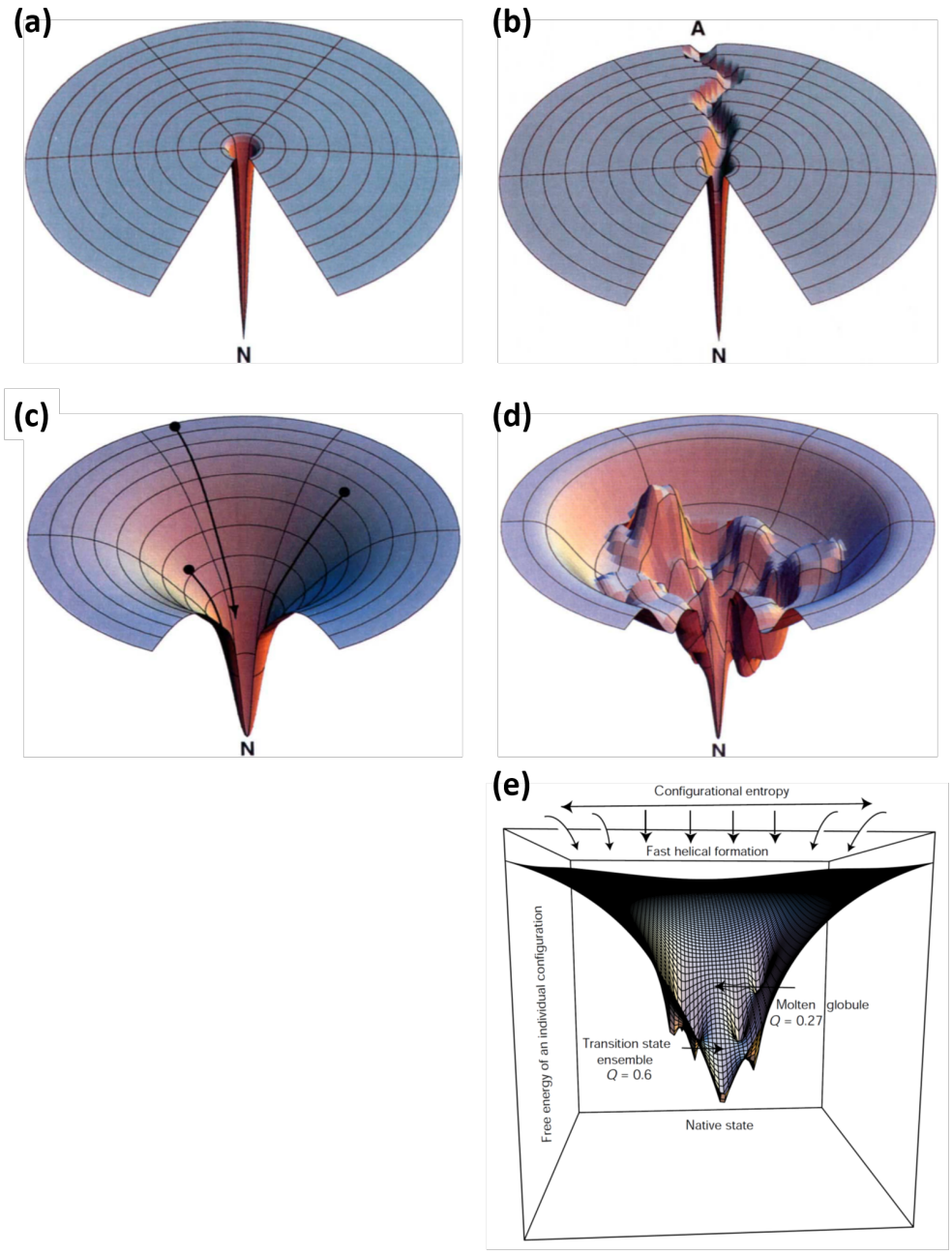


Figure 1.4: Energy landscapes
(a) The Levinthal "golf course" landscape. The polypeptide chain searches for the native conformation (N) randomly, i.e. on a flat energy landscape. **(b)** A specific folding pathway leads from a denatured conformation A to the native conformation N, hence enabling faster folding than random searching. **(c)** An idealised smooth funnel landscape. The width of the funnel represents entropy and the depth represents enthalpy. At the top of the funnel, the protein exists in many conformations with high enthalpy and high entropy. During folding, the loss of enthalpy with the formation of native-like structures compensates for the unfavourable loss of conformational entropy. **(d)** A rugged energy landscape with kinetic traps, energy barriers, and narrowing selection of paths to the native state. Panels a–d are retrieved from [53]. **(e)** Cross-section of a folding funnel from a simulation of a fast-folding helical protein. *Q* represents the fraction of correct contacts made. Retrieved from [58].

Developments in experimental and computational methods have enabled more detailed descriptions of energy landscapes. Emphasis on structural and thermodynamic parameters can be found in computational methods, including lattice models, course-grained models and atomistic simulations with implicit or explicit representations of the solvent [6,57].

In recent years, folding trajectories are studied experimentally by single-molecule fluorescence techniques. Single-molecule FRET (smFRET) data can be summarised as histograms of the native and denatured ensembles as a function of denaturant concentrations. Important thermodynamic and kinetic parameters, such as the free energy of unfolding and the interconversion rates between different states, can be derived from these histograms. smFRET studies have been performed on adenylate kinase (AK) encapsulated in lipid vesicles immobilised on a glass surface [59]. smFRET trajectories of AK show that transitions between different FRET efficiency levels are generally much smaller than the difference expected between the fully folded and the fully unfolded conformations, suggesting the presence of multiple states and that direct transitions between the folded and unfolded states are rare. Statistical analysis on the smFRET data using Hidden Markov modelling reveals six metastable states on the landscape, and enables detailed reconstruction of transition maps. The relative populations and transitions between these states vary with GdmCl concentrations, and multiple intersecting folding pathways co-exist at low GdmCl concentrations. This approach has therefore provided a comprehensive description of the folding landscape of a large multidomain protein.

Single-molecule force spectroscopy (SMFS) has also emerged as a powerful tool for reconstructing one-dimensional representations of the folding landscape based on measurements of a single molecule placed under controlled tension. Different implementations of SMFS have enabled characterisation of folding barriers and landscape ruggedness, although interpretation of reconstructed folding landscapes could be convoluted by various factors, including properties of the measurement apparatus, the presence of multiple folding pathways, and the multidimensional nature of the landscape [60].

In summary, advanced single-molecule analysis has begun to facilitate detailed characterisation of several aspects of the complex folding landscape. It is expected that future

studies will require the development of experimental methodologies with improved detection efficiencies, improved data analysis methods and the use of computer simulations in conjunction with experimental tools [61].

1.1.5 Open questions

Despite considerable advances in the soluble protein folding field, many fundamental questions remain unanswered about this complex yet fundamental biological process. Existing *in vitro* studies have often focused on simple folding systems, and the folding of more complicated systems, such as multi-domain proteins, require more in-depth studies. Purified protein samples are typically studied under equilibrium, in highly diluted and optimised conditions *in vitro*, whereas folding in the cellular milieu is more complex. Protein folding *in vivo* often begins co-translationally [62,63]. The pace of translation can regulate the vectorial folding of nascent protein domains and the protein folding landscape, thereby altering how much correctly folded, soluble protein is present in the cell. Modulation of translation rates (or even translational stalling) could occur due to various reasons [64], including the presence of rare codons in mRNAs [65–67], inefficient peptidyl transfer [68], possible presence of secondary structures in mRNAs [69–71], and interactions between positively charged residues in the nascent chain with the negatively charged ribosome exit tunnel walls [72–74]. Translation is slow (5 amino acids/second in eukaryotes or 15 amino acids/second in *E. coli*) relative to local folding events (μs–ms timescale) [75]. Vectorial folding in the confined space of the exit tunnel and interactions between the nascent chain and the exit tunnel are likely to define the landscape of protein folding, and result in lower energies than would be the case for protein synthesis without co-translational folding [62]. Given the differences between *in vivo* and *in vitro* folding landscapes, co-translational folding intermediates may also be structurally dissimilar to *in vitro* refolding intermediates [76–78].

The ribosome also serves as a platform for protein biogenesis factors to interact with the emerging nascent chain and effect its maturation and proper cellular localisation. Emerging peptides can first interact with the ribosome surface [79–81], during which the ribosome may exert a chaperoning effect, protecting the nascent chain from misfolding and aggregation until the protein is fully synthesised and extruded from the peptide exit tunnel [64,77]. Beginning with N-terminal processing, the nascent polypeptide chain emerging from the

ribosome has access to a huge ensemble of chaperones and quality control mechanisms to assist its folding, cellular localisation and proteastasis [82,83]. Thus, in contrast to the utilisation of purified protein samples in *in vitro* experiments, it is unclear how much time proteins spend unfolded and unprotected *in vivo*.

Moreover, protein folding *in vivo* occurs in spatially organised (e.g. for membrane proteins and in eukaryotic cells), inhomogeneous and crowded environments with macromolecule concentrations ranging from 200 to 400 mg/ml. These properties are often very difficult to mimic *in vitro* [75].

1.2 Folding of membrane proteins

Folding of integral membrane proteins adhere to similar physical principles to those for soluble protein folding. Integral membrane proteins are constrained by the additional need to compensate for the energetic cost of ~1.2 kcal/mol per peptide bond for burying peptide bonds in the lipid bilayer. This high cost of interfacial partitioning of the peptide bond could explain the coupled insertion-folding mechanisms and the tendency of membrane spanning regions to form regular secondary structural elements in order to maximise the hydrogen bonding potential. Secondary structure formation reduces the energetic cost of incorporating peptide bonds into a bilayer by ~0.4 kcal/mol per peptide bond for α-helical structures and ~0.5 kcal/mol per peptide bond for β-sheet structures [84]. Hence, although the lipid bilayer is a complex and inhomogeneous environment, it imposes thermodynamic and geometric constraints which limit the number of feasible structural motifs to α-helical bundles and β-barrels.

Membrane protein misfolding is implicated in the progression of various diseases such as cystic fibrosis, Charcot-Marie-Tooth disease and retinitis pigmentosa [85]. Insights on the folding mechanisms of membrane proteins might pave the way towards discovering new therapeutic approaches for membrane misfolding diseases.

Advances in studying membrane protein folding progress slower than that for globular protein folding due to various experimental challenges [3]. Characterisation of the denatured states, and the unfolding and folding pathways require maximal and reversible unfolding of the protein with minimal aggregation. *In vitro* studies require membrane proteins to be functionally reconstituted in a detergent or lipid environment, which often also renders them resistant toward denaturation. The inability of membrane proteins to unfold completely in a lipid or detergent environment is related to water being less available to hydrogen-bond with protein backbone and side-chain moieties in the membrane [86]. Another challenge is the formation of aggregates due to the exposure of large hydrophobic regions upon membrane protein denaturation. Moreover, identifying experimentally favourable folding conditions for any particular membrane protein is a difficult process of finding the suitable combination of detergent or lipid composition, buffer, pH and temperature [18]. Similar to the soluble protein

folding field, chemical denaturants, pH and temperature have been used as denaturants in membrane protein unfolding/folding studies [18]. Several studies indicate the limited use of temperature in deriving thermodynamic parameters for both α-helical and β-barrel membrane proteins as thermal denaturation of membrane proteins is often irreversible. Like for soluble proteins, the choice of denaturant and the membrane protein of study both profoundly affect the conformations represented in the denatured states ensemble.

Unfolding/folding studies on membrane proteins have historically focused on bacteriorhodopsin [87] and OmpA [88]. The studies to date tend to be dominated by proteins of bacterial or archaeal origin, with only a handful of eukaryotic proteins having been explored [18]. Several other membrane proteins have also emerged as model systems which are amenable to detailed kinetics and thermodynamics studies, leading to the proposal of different folding models for β-barrel and α-helical membrane proteins.

1.2.1 β-barrel

β-barrel membrane proteins are formed by membrane-spanning β-strands which are held together by a lateral hydrogen bond network, resulting in an overall cylindrical topology in order to satisfy all possible hydrogen bonds in the peptide backbone. In contrast to the ubiquitous distribution of α-helical transmembrane proteins, β-barrel membrane proteins are found exclusively in the outer membranes of Gram-negative bacteria, mitochondria and chloroplasts [89].

β-barrel transmembrane proteins offer distinct technical advantages over their α-helical counterparts, as β-barrel transmembrane proteins have much smaller average sequence hydrophobicities and can be completely unfolded in aqueous denaturants and refolding into membranes can be studied from such states [90]. While aqueous denaturants such as urea and GdmCl generally induce the canonical unfolded, random coil state, these denatured states can still contain measurable amounts of residual structure. For example, an NMR study of the denatured state ensemble of outer membrane protein X (OmpX) shows the presence of non-random structure in urea [91].

The general mechanism for β-barrel membrane protein membrane insertion and folding has been elucidated through time-resolved tryptophan fluorescence quenching (TDFQ) experiments on OmpA [89,92]. Tryptophan fluorescence quenchers are placed at different depths in the membrane, and the time courses for the passage of tryptophans past these zones of quenchers are followed. All tryptophan residues of OmpA cross the membrane with the same timecourse, suggesting that individual transmembrane β-strands are translocated in a concerted fashion. Different membrane insertion and folding intermediates of OmpA were also characterised by TDFQ. The five tryptophan residues which are distributed across four transmembrane segments of OmpA have specific distances from the center of the bilayer in each of the folding intermediates (Figure 1.5) [93]. The unfolded state first hydrophobically collapses into a water-soluble intermediate, which then binds to the membrane surface to form the first membrane-bound intermediate with disordered tryptophans located 14–16 Å from the center of the bilayer. This intermediate then forms a molten disk structure with some developed β structure and all tryptophans located at ~ 10 Å away from the bilayer center. Next, the four tryptophans on the four β-hairpin loops translocate through the bilayer while forming a molten globule intermediate. The molten globule intermediate then proceeds to the native state, in which all tryptophans are located at ~ 10 Å from the bilayer center. Identical kinetics were observed for secondary structure (CD spectroscopy) and tertiary structure (cold SDS-PAGE) formation, indicating that inter-strand hydrogen bonds and the barrel structure form while the protein translocates across the membrane [94]. The co-translocational folding mechanism is logical for β-barrel membrane proteins, because membrane insertion after barrel formation would have required the creation of a large defect in the membrane to insert the fully-formed barrel. Further studies have involved characterisation of the dependence of the free energy of unfolding and the *m*-value on the thickness of phosphatidylcholine bilayers, showing that folding into thick bilayers is a thermodynamic two-state process whereas folding into thin bilayers is a multi-state process [95]. These studies indicate that increasing the lateral bilayer pressure increases the thermodynamic stability of the protein, whereas hydrophobic mismatch leads to easier penetration of water molecules into the hydrophobic core of thin bilayers and stabilisation of folding intermediates.

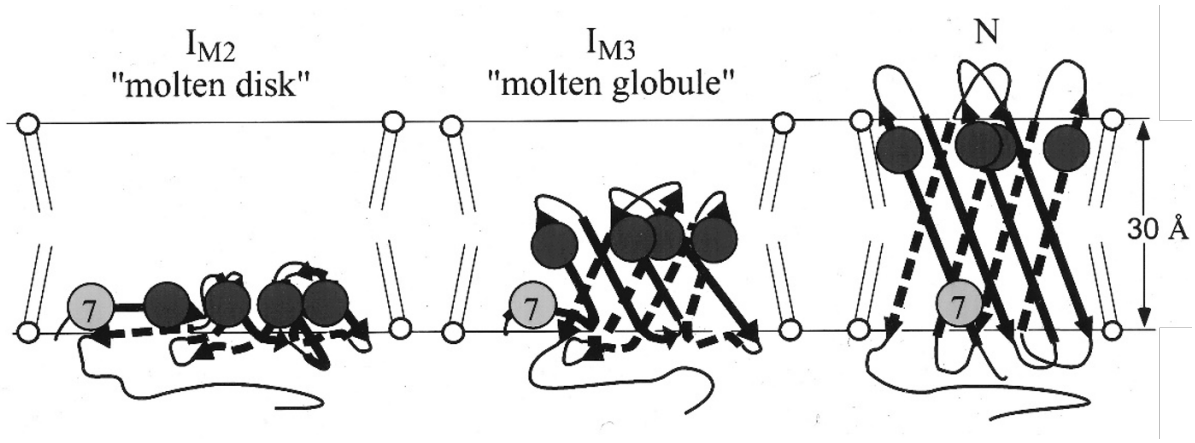


Figure 1.5: Schematic for the concerted folding and insertion mechanism of OmpA
The collapsed OmpA first adsorbs to the membrane surface, with all Trps located on the *cis* monolayer, at 14–16 Å away from the center of the bilayer (not shown). An I_{M2} "molten disk" intermediate forms with all Trps at 10 Å away from the bilayer center. Trp-15, Trp-57, Trp-102 and Trp-143 then translocate through the bilayer while forming the I_{M3} "molten globule" intermediate. Finally, the native state (N) is formed, with all Trps located 10 Å away from the bilayer center, and only Trp-7 remaining on the *cis* monolayer. This model shows nearly synchronous insertion and translocation of Trp-15, Trp-57, Trp-102 and Trp-143 across the bilayer. Figure adapted from [93].

One of the major unresolved questions is how folding models derived from *in vitro* studies can be reconciled with the transport and insertion of β-barrel membrane proteins *in vivo*, especially considering the requirement for β-barrel membrane protein precursors to traverse multiple membranes with different lipid and protein compositions during cellular localisation followed by chaperone-assisted folding and insertion in different non-cytosolic milieus. *In vivo*, bacterial β-barrel membrane proteins are synthesised in the cytosol, secreted by the SEC translocase across the inner membrane into the bacterial periplasm and the unfolded proteins are subsequently inserted and folded into the Gram-negative outer membrane with the assistance of the bacterial β-barrel machinery (BAM) [96–98]. In the bacterial periplasm, Skp [99] and SurA (a peptidylprolyl isomerase) [100] assist the folding of several membrane proteins including OmpA, OmpF and LamB [90] and act as chaperones to prevent the aggregation of precursors. In eukaryotic mitochondria, β-barrel precursor proteins synthesised in the cytosol are translocated through the translocase of the outer membrane (TOM) into the intermembrane space [101]. Small TIM complexes act as chaperones in the intermembrane space [98]. X-ray crystallography and NMR studies suggest that oxidised small TIM monomers assemble into hexameric chaperone complexes [102,103] in the intermembrane space and hold client proteins in a nascent chain-like extended conformation

to facilitate insertion into the mitochondrial inner or outer membrane [104]. β-barrel membrane proteins are inserted into the mitochondrial outer membrane by the sorting and assembly machinery at the outer membrane (SAM/TOB). α-helical metabolite carrier proteins are inserted into the mitochondrial inner membrane by the carrier translocase of the inner membrane (TIM22) in a membrane potential-dependent manner [96–98].

1.2.2 α-helical

α-helical membrane proteins are found in most eukaryotic membranes and in the inner membranes of bacteria [18]. Studies on denaturation and refolding of α-helical membrane proteins began with research from the Khorana group, showing that bacteriorhodopsin can be renatured from a fully-denatured state in 88% formic acid or anhydrous trifluoroacetic acid via an SDS-solubilised state [87]. Kinetic folding studies on α-helical membrane proteins typically begin with a partially, reversibly denatured state. Most α-helical membrane proteins are resistant to aqueous denaturants, with a few exceptions including the denatured state ensembles of diacylglycerol kinase (DAGK) in GdmCl or urea in the presence of 1% formic acid [105], and tobacco light-harvesting chlorophyll a/b protein complex (LHCIIb) in neutral GdmCl [106]. Mixed micelles containing ionic detergents such as SDS have been shown to be efficient denaturants for some cases, though the proteins are likely to be recalcitrant to complete unfolding, possibly because SDS tends to favour the formation of non-native helical structures [18]. The SDS-denatured bacterio-opsin apo-protein shows partial loss of secondary structure and loss of retinal binding, and can be renatured upon the addition of renaturing phospholipid/cholate mixtures [107]. Folding studies are typically initiated by mixing denatured protein with renaturing detergents or lipids, and followed by changes in intrinsic fluorescence, secondary structure and recovery of protein function. Reversible chemical denaturation has only been demonstrated for a handful of α-helical membrane proteins [87,105,108–115]. Nevertheless, biophysical, thermodynamic and kinetic studies have enabled extensive characterisation of the molecular nature of denatured states [87,105–107,109,111,114,115], protein stability [108,110,111,115,116], transition state [117,118], and folding intermediates [119,120], hence yielding valuable insights on the folding mechanisms of α-helical membrane proteins.

Very different folding mechanisms have been proposed for α-helical vs. β-barrel transmembrane proteins. This is likely due to their distinct hydrogen bonding patterns. A widely held concept for explaining the folding of integral α-helical membrane proteins is the two-stage model originally proposed by Popot and Engelman [86] (Figure 1.6a). This model postulates that transmembrane helices are first formed independently after insertion into the membrane environment, followed by formation of tertiary packing during the second stage. Parallels can be drawn between the two-stage model for α-helical membrane protein folding and the framework model for soluble protein folding. The essence of the two-stage model is that much of α-helical membrane protein folding can be understood by focusing on helix association in the bilayer without substantial consideration of the insertion process. Though possibly oversimplified, the two-stage model nevertheless led to subsequent informative studies on helix-helix interactions, including identification of the GxxxG motif for helix dimerisation [121], the importance of stereochemical fit between side-chains of helices in interhelical interactions [122], and the role of hydrogen bonds between main chain Cα–H and backbone or side-chain oxygen atoms [123].

Bacteriorhodopsin is a light-driven proton pump in the purple membrane of *Halobacterium salinarium*. The SDS-denatured apo-protein can be spontaneously refolded to the native state [107], enabling *in vitro* thermodynamics and kinetics studies on folding and unfolding. The two-stage model originated from successful reconstitution of the native chromophore in bacteriorhodopsin by assembling complementary combinations of proteolytic fragments [124,125] and/or complementary synthetic peptides [126,127], or complementary/overlapping fragments which have been individually expressed in and purified from *E. coli* [128]. Absence of one or more of the transmembrane segments results in abolishment of pigment formation. These results indicate that the assembly of bacteriorhodopsin requires each of the seven helices, whereas the surface loops appear to be dispensable. The large retention of secondary structure in the SDS-denatured state of bacteriorhodopsin also appears to support the idea that individual secondary structure elements can be stabilised in a micellar environment and are formed before native tertiary packing is established. However, the notion of individual transmembrane helices folding independently in a membrane environment has been partly questioned by structural studies on isolated helical fragments of bacteriorhodopsin, which show that C-terminal helices are

only partially helical and require the presence of N-terminal helices to fold [129], hence implying that tertiary contacts are necessary for some transmembrane helices to acquire stable folds.

The toolkits for studying soluble protein folding have been successfully applied to *in vitro* kinetics studies of bacteriorhodopsin. SDS-denatured bacteriorhodopsin, with covalently-bound retinal and absorption peak at 440 nm (bO_{440}), folds with two-state kinetics. The transition state is much closer to the denatured state than to the native state ($\beta_T \sim 0.14$). ϕ-value analyses on helices B [117] and G [130] by the Booth group reveal high ϕ-values ($\phi > 0.8$) in helix B, suggesting that this helix is structured in the transition state, thus forming a folding nucleus for the rest of the protein. Packing and consolidation of helix G occurs late in folding, with low ($\phi \leq 0.2$) or intermediate ($0.3 < \phi < 0.7$) ϕ-values being observed. Similar studies pursued by Bowie *et al*. [118], however, show consistently low ϕ-values of less than 0.4 across the protein, suggesting that the transition state is a loosely organised ensemble of conformations. It was therefore proposed that the rate-limiting step of refolding is the topological rearrangement of largely preformed helices from an ensemble of rapidly interconverting conformations. Differences in the results between the studies by the Booth and Bowie groups might be attributable to the different detergent/lipid conditions used in the two studies. Nevertheless, interpretations of ϕ-values are rather reserved in both studies, given that the SDS-denatured bO_{440} retains substantial α-helical content and the main structural changes in refolding are the formation of native interhelical packing and tertiary structure.

The two-stage model was subsequently further developed by adding an additional stage to account for ligand binding after helix association [131]. Retinal binding in SDS-denatured bacterioopsin has been characterised on a ms–min timescale using circular dichroism [132], stopped-flow fluorescence spectroscopy [120] and absorption spectroscopy [133], leading to the proposal of the following reaction scheme [134]:

$$bO \rightleftharpoons I_1 \rightleftharpoons I_2 \overset{R}{\rightleftharpoons} I_R \longrightarrow bR$$

SDS-denatured apo-bacterioopsin (bO) refolds spontaneously upon diluting SDS with renaturing micelles or vesicles. Two intermediates, I_1 and I_2, are formed prior to retinal

binding. The formation of the transient I_1 intermediate is accompanied by a change in light scattering of the micelles and could reflect a change in the micelle structure or protein folding. I_2 is a key apoprotein intermediate with native helical content. Its formation is the rate-limiting step in apo-protein folding, and the rate can be controlled by altering lipid composition or mutations. Different I_2 states can be observed depending on the folding conditions. In the absence of retinal, an equilibrium is established between bO and I_2. Retinal (R) binding occurs in at least two steps, first non-covalently to yield I_R followed by formation of a covalent Schiff base linkage to Lys-216 to yield the native chromophore. I_R consists of at least two observable states which absorb at 380 and 440 nm, respectively, both of which form bacteriorhodopsin (bR) with absorption at 560 nm at the same rate. This folded state initially contains covalently-bound all-*trans* retinal, which then isomerises over the course of about 1 h to yield a mixture of all-*trans* and 13-*cis* retinal, analogous to the representation of retinal isomers in the dark-adapted state [134]. The retinal cofactor is therefore a major contributor to the unfolding and folding kinetics of bacteriorhodopsin.

The folding of rhodopsin deviates substantially from the two-stage model. Rhodopsin is a G protein-coupled receptor which functions as a dim light photoreceptor in the retina. Rhodopsin shares a similar transmembrane heptahelical fold but no sequence similarity to bacteriorhodopsin. SDS-denatured rhodopsin reveals extensive abolishment of helicity and native tertiary interactions, and the formation of largely unfolded structures with some residual structure elements causing the burial of specific cysteines and tryptophans [135]. Cell-based [136] and *in vitro* [135,137,138] folding studies on rhodopsin suggest that interactions between the extracellular and transmembrane domains constitute a “folding core” which is important in the early stages of folding. Single-molecule dynamic force spectroscopy reveals a core of rigid structural elements in bovine rhodopsin but not in bacteriorhodopsin [137]. NMR and EPR studies of SDS-denatured rhodopsin show flexible cytoplasmic and extracellular ends while some residual structure remains in the transmembrane and extracellular regions [138]. The distributions of rigid and flexible segments in rhodopsin identified in these experiments are consistent with the results from simulated thermal unfolding studies using the Floppy Inclusions and Rigid Substructure Topography (FIRST) method. In comparison, simulation with bacteriorhodopsin shows clusters that align with individual helices and break off early in the simulation. The unfolding

mechanism of bacteriorhodopsin therefore illustrates better consistency with the two-stage model [139]. The more recent "long-range interactions model" [140] (Figure 1.6b) has been proposed in order to emphasise the importance of interactions between residues which are far apart in sequence but come close in space during early stages of protein folding. The long-range interactions model is similar to the nucleation condensation model for soluble proteins. To date, rhodopsin is the only α-helical membrane protein which is known to adhere the long-range interactions model. Detailed comparison of the modes of activation and signal transduction across different membrane receptors underline the importance of all three types of long-range interactions: (1) between transmembrane helices; (2) between helices and soluble loops; and (3) between loops [140]. Fundamental understanding on the roles of such interactions in the folding, stability and function of membrane proteins will have profound implications on addressing membrane protein misfolding diseases.

Unfolding and refolding of disulphide bond reducing protein B (DsbB) in SDS and DDM have been fitted to a three-state folding scheme [141]. Three linear regimes are observed in the chevron plot of DsbB, at 0–0.2 X_{SDS}, 0.3–0.5 X_{SDS} and > 0.5 X_{SDS}, respectively. The chevron plot is consistent with the presence of an intermediate state I which accumulates at high X_{SDS}, and a rate-limiting transition state between I and denatured state D. ϕ-value analysis of alanine-scanning mutants of DsbB reveals a folding nucleus in the transition state consisting of 2–3 residues with $\phi > 0.5$ clustered at one end of the transmembrane helical bundle. In the intermediate state, the nucleus expands to include residues closer to the middle of the protein. This suggests early formation of structure near the tips of the transmembrane helices which nucleate subsequent structure formation in other regions of the protein.

To summarise, seminal denaturation, kinetics and thermodynamics studies on bacteriorhodopsin have by-and-large supported the two-stage model for α-helical membrane protein folding, although some alterations have been recommended in view of structural studies on isolated helices, suggesting some importance of tertiary interactions. Most α-helical membrane proteins which have been studied to date appear to adhere the two-stage model. While alternative models have been postulated, including the long-range interactions model for rhodopsin and the three-state model for DsbB, studies with new model proteins and

across comparable experimental conditions are required to further explore these alternative models for α-helical membrane protein folding.

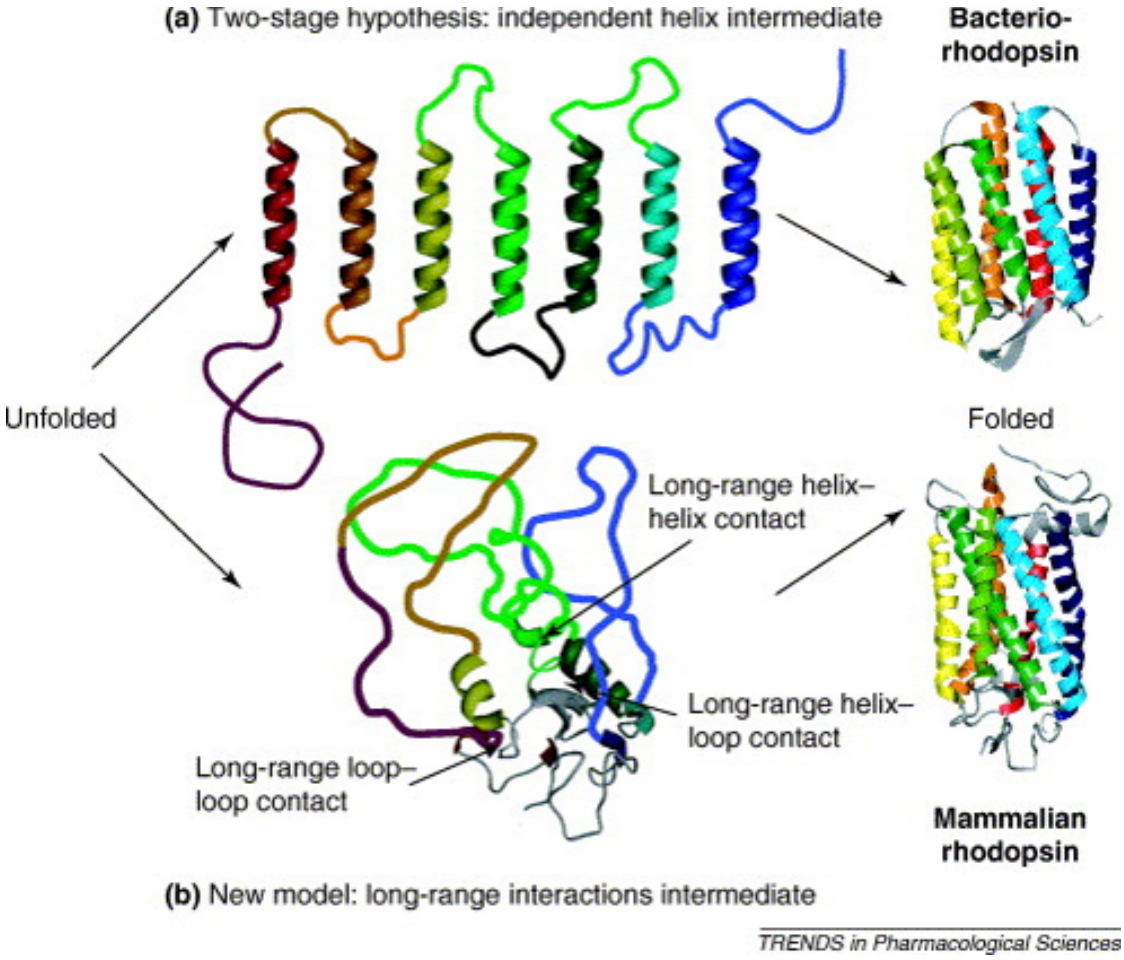


Figure 1.6: Models for α-helical membrane protein folding
(a) The two-stage model postulates that transmembrane helices are first formed independently after insertion into the membrane environment, followed by formation of tertiary packing. **(b)** The long-range interactions model emphasises the importance of long-range tertiary contacts in the formation of a "folding core" during early stages of protein folding. The "folding core" is identified as elements of residual structure in the denatured states, and provides a scaffold for the folding of the rest of the protein. Figure retrieved from [140].

Most prokaryotic and eukaryotic α-helical membrane proteins are co-translationally inserted into the membrane *in vivo* via the SecYEG/Sec61 translocon [142,143] or YidC-like insertases [144,145]. In the Sec pathway, the first hydrophobic element, either a cleavable N-terminal signal sequence or a transmembrane domain, emerges from a translating ribosome, and is recognised by the signal recognition particle (SRP) [146]. SRP and its receptor, SR (or FtsY), form a heterodimeric complex, which stimulates GTP hydrolysis on both proteins,

thereby enabling transfer of the SRP/FtsY/nascent chain ternary complex to the translocon [147]. The ribosome exit tunnel is located in close proximity to the translocon pore, so that the nascent chain emerges from the ribosome directly into the translocon channel. The transmembrane helices partition into the membrane via a lateral gate in the translocon [143]. The association of SecY with the ribosome provides the energy for co-translational insertion. Integral membrane proteins with large hydrophilic periplasmic domains also require the presence of a cytoplasmic SecA ATPase to translocate such domains across the membrane [143,148]. An alternative model suggests that the transmembrane helices do not enter the Sec translocon. Instead, transmembrane helices first partition with the interfacial headgroup region of the lipid bilayer before inserting into the membrane by sliding down the lateral gate of the translocon. In this model, the helices are constantly in contact with the lipid bilayer, and membrane insertion is driven by the highly favourable association of the membrane protein with the surrounding lipids [149].

Proteins can also be inserted into membranes via YidC-like insertases, including YidC from *E. coli*, Oxa1 from mitochondria, and Alb3 and Alb4 from chloroplast thylakoids [145]. The mechanisms of YidC-like insertases are less well understood. Unlike the Sec translocon, YidC does not possess a membrane-spanning pore or channel. It has been proposed that the hydrophilic region of the membrane protein substrate is translocated across the periplasmic membrane by first binding to the hydrophilic groove of YidC, while the hydrophobic transmembrane segments of the membrane protein slide along the transmembrane region of YidC and into the lipid bilayer [144,150]. The currently identified range of substrates for the YidC-only pathway is limited, but are strikingly all part of large oligomeric assemblies, thus leading to speculations that YidC is involved in the assembly of multimeric membrane protein complexes [151]. YidC can also form a functional complex with the Sec translocon, and interact with transmembrane helices as they emerge from the lateral gate of SecYEG, although the functional importance of this interaction is not well understood [145].

Various factors affect the efficiency of transmembrane insertion and folding *in vivo*. Gating of the Sec translocon and membrane insertion are determined by the average hydrophobicity of transmembrane helices, the length of a given transmembrane helix, its amino acid composition, and the relative positions of amino acids within the transmembrane domain

[143,152,153]. The Sec translocon and YidC insertase have been proposed to have important roles in membrane protein folding. In certain SecY mutants, lactose permease LacY can be inserted but is misfolded and susceptible to cellular proteolysis [154]. Susceptibility to misfolding and proteolysis of LacY are also observed when YidC is depleted [155]. These data suggest that SecY and YidC are important in the insertion and folding of membrane proteins. Interactions between the translocon and the substrate membrane protein could facilitate early folding events. A study on the insertion of aquaporin-4 (AQP4) water channel via the Sec61 translocon shows that each transmembrane helix of AQP4 moves through the translocon in a highly ordered and sequential fashion, but up to four transmembrane helices of AQP4 can simultaneously contact Sec61α after being displaced from their initial site of interaction. This suggests that the translocon uses a single primary site for transmembrane insertion while providing a scaffold or protected environment for facilitating early protein folding [156]. The lipid bilayer can also influence membrane protein insertion [157–161], folding [110,162–164], topology [165,166], oligomerisation [167,168] and function [169–172], either directly through specific interactions with the headgroup region, or indirectly via the lateral pressure of the bilayer [173]. Moreover, protein-lipid interactions are crucial during various stages of membrane protein insertion and folding, including recognition of the inserting transmembrane segment by the translocon [174,175], the release of nascent chains from the SRP-FtsY complex [176–178], and stimulation of the GTPase activity of FtsY [179,180]. Given the large variations in the lipid compositions of native membranes between different organisms and organelles [181], such complex native membrane environments are difficult to replicate in *in vitro* experiments [173].

Summarising the above, membrane protein insertion and folding *in vivo* is complex and dynamic. While the majority of work on α-helical membrane protein folding has so far been done *in vitro* and has yielded valuable kinetic and thermodynamic information, reconciling *in vitro* reversible folding studies with co-translational folding *in vivo* remains as a very challenging goal in this field.

1.3 Hydrogen bonding

A hydrogen bond (H-bond), typically represented as X−H···Y−Z, is an attractive interaction between a hydrogen atom from a molecule X−H, in which X is more electronegative than H, and atom(s) in the same or different molecule in which there is evidence of bond formation (Y−Z) [182]. H-bonds have an intermediate energy range (0.5−40 kcal mol^{-1} per H-bond) that spans between the energy levels for van der Waals interactions and covalent interactions. This explains the functional versatility of H-bonds both in stabilising static structures and in mediating dynamical processes [183].

H-bonds underpin protein stability, structure and function. The hydrophobic environment of the lipid membrane necessitates membrane proteins to establish hydrogen bonds to avoid exposing their polar groups to the lipid environment, which would be energetically highly unfavourable. All H-bonds in these regions of the membrane protein must be self-satisfied [184]. In α-helical membrane proteins, backbone H-bonds are aligned parallel to helical axes and stabilise the individual helices. Backbone H-bonds in β-barrel proteins are formed between neighbouring β-strands and stabilise the tertiary structure.

1.3.1 Methods for studying H-bonds

Accurate description of H-bonds has been a long-standing goal for experimental and theoretical studies in structural biology. Knowledge about the geometric properties, including distance and torsion angles, dynamics and strength of H-bonds, is essential for further understanding of the highly context-dependent contribution of H-bonds to protein stability, structure and function. It is difficult to obtain precise information on H-bonds from X-ray crystallography due to the weak scattering density of the hydrogen atom. The presence of H-bonds is usually inferred from the spatial arrangement of the donor and acceptor groups in crystal structures. Generally, a H-bond is assumed to be present when the distance between the donor and acceptor atoms is less than or equal to the sum of the van der Waals radii of the two electronegative atoms [182]. High-resolution solution-state NMR has contributed significantly towards the understanding of H-bonds. A number of NMR observables provide indirect evidence for H-bonds, including kinetic hydrogen/deuterium exchange, the equilibrium hydrogen/deuterium isotope fractionation factor, changes in chemical shift, scalar

coupling and deuterium quadrupole coupling [185]. Characterising structure-to-function relationships of H-bonds by solution-state NMR, nevertheless, remains challenging due to the weak strength, dynamic nature and abundance of H-bonds in bio-macromolecules.

Many polar hydrogen atoms, such as those of amide, imino and hydroxyl groups, continuously exchange with solvent hydrogens. This exchange is significantly slowed down when the hydrogen participates in an H-bond, even when this bond is directly exposed to the solvent. The rate of hydrogen/deuterium exchange (k_{ex}) is described as a combination of the rates of the opening (k_{op}) and closing (k_{cl}) of the hydrogen bond in addition to the inherent rate of exchange of the hydrogen once it is in the open state (k_{int}) [186]. For an amide group:

$$\mathrm{NH_{cl}} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \mathrm{NH_{op}} \xrightarrow{k_{int}} \mathrm{ND} \tag{1.23}$$

Once an amide hydrogen reaches the open state, the exchange kinetics is defined by the balance between reverting to the closed state and moving forward with exchange.

H/D exchange kinetics detected using NMR spectroscopy and mass spectrometry have produced numerous interesting studies in the context of protein folding. For example, residue-specific kinetics of inter-strand H-bond formation has been studied in the β-barrel outer membrane protein X (OmpX) in detergent micelles, showing that the kinetics of H-bond formation are uniform across the entire β-barrel and synchronised with tertiary structure formation. This suggests that OmpX folding progresses via an ensemble of different short-lived, partially-folded and interconverting conformations in which all backbone amides are accessible to solvent exchange and engage in H-bonds transiently [187].

H-bond strength is traditionally examined by disrupting native H-bonds through chemical (amide-to-ester or amide-to-*E*-olefin) [188] or side-chain [189] mutagenesis, hence limiting such investigations to only a few bonds at a time and possibly also leading to disruption of the native protein structure. H/D fractionation factors (Φ) have been widely used to characterise H-bond strengths in small molecules and proteins.

$$\mathrm{NH_{protein}} + \mathrm{D_{solvent}} = \mathrm{ND_{protein}} + \mathrm{H_{solvent}} \tag{1.24}$$

$$\Phi = \frac{([\mathrm{ND}]/[\mathrm{NH}])_{\mathrm{protein}}}{([\mathrm{D}]/[\mathrm{H}])_{\mathrm{solvent}}} \tag{1.25}$$

Φ is determined by the difference in vibrational energy between N–H and N–D bonds at an exchangeable site [190]. Due to its larger mass, the vibrational energy minimum (or zero-point energy) for deuterium is lower than that for a proton. This principle applies to both solvent and protein. If the energy difference is larger in a protein amide site than in the solvent, the amide site will be enriched in deuterium (Φ > 1). Conversely, Φ < 1 indicates a preference for proton over deuterium. Studies on small soluble proteins [191–194] show that most Φ values are close to 1, and that Φ < 1 have been observed for "strong" H-bonds, whereas amide sites with H-bonds which are weak compared to solvent–solvent H-bonds have Φ > 1.

H/D amide isotope substitution has provided an avenue for identifying the formation of H-bonds during protein folding. Equilibrium and kinetic isotope effects are quantified by measuring the change in stability due to replacement of deuterated amides with hydrogens from the solvent, $\Delta\Delta G^{D-H}$. Isotope substitution of backbone amide leads to changes in denaturant dependence of folding and unfolding rates. The perturbation of folding free energy $\left(\Delta\Delta G_f^{\ddagger D-H}\right)$ can then be compared against the global stability change ($\Delta\Delta G^{D-H}$) using an isotopic ϕ_f^{D-H} value:

$$\phi_f^{D-H} = {\Delta\Delta G_f^{\ddagger D-H}}\Big/{\Delta\Delta G^{D-H}} \tag{1.26}$$

This method has been applied to characterising the extent of H-bond formation in the transition state of soluble proteins which fold with two-state kinetics. Cross-linked and dimeric GCN4-D7A have ϕ_f^{D-H} of ~0.59, indicating that about half of the helical H-bonds are formed in the transition states of the coiled-coil [195]. While $\Delta\Delta G^{D-H}$ is found to correlate with the number of helical hydrogen bonds, β-sheet and long-range hydrogen bonds have little isotope effect [196]. Hence, there is limited general applicability of H/D amide isotope substitution to different structural motifs.

The application of H/D exchange-based methods to studying H-bonds relies fundamentally on detailed quantification of the amount (at equilibrium) and the rate of solvent exchange within the suitable experimental timescales. Membrane proteins often have little access to water molecules when partitioned in the detergent or bilayer milieu for *in vitro* studies.

Measurement and interpretation of kinetic and equilibrium thermodynamic parameters derived from H/D exchange could therefore be complicated by their limited accessibility to the bulk solvent.

Chemical shift changes are usually observed for nuclei involved in H-bonds, as H-bond formation leads to redistribution of electron densities. Various trends have been noted, for example H-bonding to an electronegative acceptor atom leads to a downfield shift of the isotropic ^{1}H chemical shift, while H-bonding to π-electrons in aromatic rings usually results in an upfield shift [185]. However, the correlation between changes in H-bonding and chemical shift changes could often be convoluted by other factors, such as structural changes in neighbouring or distal residues.

Quadrupole coupling refers to the interaction of a quadrupolar nucleus with an electric field gradient. The electric field gradient arises from an asymmetric distribution of electron density around the nucleus. The quadrupole coupling constant (QCC) of a deuteron within an H-bond is very sensitive to the asymmetric charge distribution in the H-bond [185]. The T_1 relaxation rate of a deuteron is dominated by its QCC. QCCs for backbone amide deuterons in human ubiquitin show good correlation with H-bond distances and angles in the crystal structure [197]. Precise measurement of QCC requires accurate structural knowledge for modelling the rotational diffusion of individual N–D bond vectors and accurate measurement of the $^{1}J_{ND}$ scalar coupling constant. The main challenges with accurate measurement of QCC lies with the broad linewidths of quadrupolar nuclei due to fast relaxation. Hence, this methodology may be somewhat restricted to proteins with low molecular weight and dynamics which are favourable on the NMR timescale [197].

1.3.2 Scalar coupling

Scalar coupling (or *J*-coupling) arises from the magnetic interaction between nuclei and is mediated by bonding electrons [185,198]. The nuclear spin polarisation of one atom affects the polarisation of the surrounding electrons (Figure 1.7a). The polarised electron cloud produces a change in the magnetic field surrounding the coupled spin. For example, in amides, the bonding electrons in the σ orbital become polarised by the proton magnetic dipole. Protons have two spin states. Depending on the orientation of the proton spin, the

polarisation of electrons can either increase or decrease the magnetic field at the N nucleus, causing either an upfield or a downfield shift in the resonance line. The observed spectrum thus contains two peaks which are separated by the *J*-coupling constant (resonance splitting) (Figure 1.7b). *J*-couplings are denoted as ${}^{n}J_{AX}$, where A and X are the interacting nuclei, and *n* is the number of interceding covalent bonds. The size of the coupling constant depends on the number of bonds separating the coupled nuclei, and on the configuration and spatial arrangements of the bonding electrons. The value of the coupling constant can be either positive or negative.

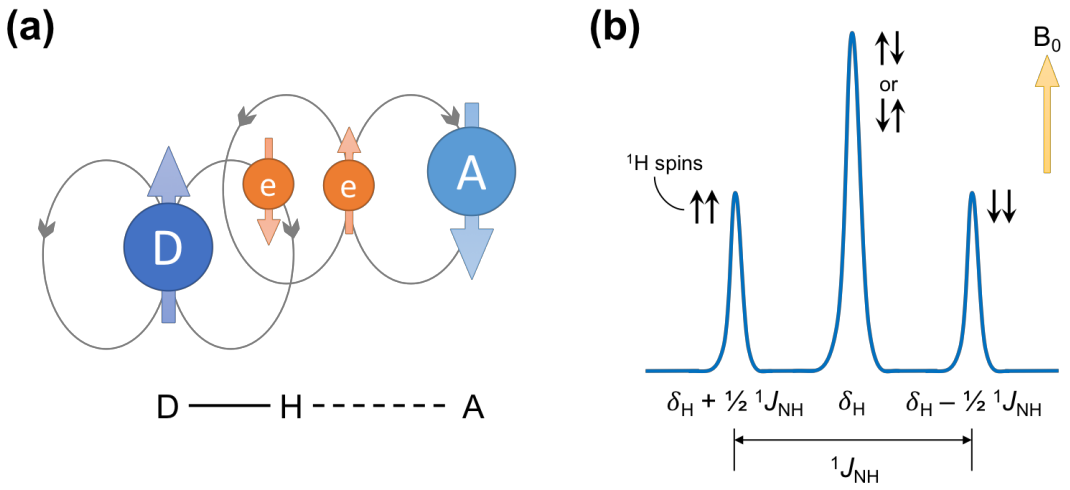


Figure 1.7: Scalar coupling
(a) Schematic representation of the scalar coupling mechanism across H-bonds. **(b)** Schematic diagram showing the splitting of an amide ^{1}H resonance by the $^{1}J_{NH}$ coupling constant. The direction of the applied magnetic field (B_0) is shown by the yellow arrow. The orientations of the ^{1}H spins for different resonances are shown using black arrows. Figure adapted from [185].

Since *J*-couplings are mediated by bonding electrons, they yield valuable information on the chemical connectivity of atoms and structures of bio-macromolecules, and provide angular restraints in NMR structure calculations [198]. *J*-coupling also underpins the transfer of magnetisation across bonded systems in multidimensional NMR [199]. The magnitude of *J*-couplings should always be compared against the linewidth of associated signals. Since linewidth increases with molecular weight, the detection of small, long-range *n*-bond couplings would be difficult for large proteins [198]. The typical values [198,200] and applications of several *J*-coupling constants in proteins are summarised in Table 1.1.

Previous investigations have elucidated the $^1J_{NH}$ values in the random coil protein Tau, thereby providing an amino acid-specific reference for non-H-bonded residues. Xiang *et al*. have identified H-bonds by calculating Δ^1J_{NH}, the difference of $^1J_{NH}$ values measured in folded proteins (e.g. ubiquitin) and the amino acid-specific random coil values [201]. The authors have provided a modest error range of ± 0.35 Hz to account for perturbations in the random coil $^1J_{NH}$ measurements due to differences in pH and magnetic field strength between their different samples. In ubiquitin, residues with Δ^1J_{NH} smaller (i.e. more negative) than −0.35 Hz correlate well with H-bonding predictions by the WhatIf server [202].

$^nJ_{AX}$	*J* (Hz)	Applications & Empirical relationships		Ref.
$^1J_{NH}$	(–)89–95	Presence of backbone H-bonding: H-bond is detected as Δ^1J_{NH}, the deviation of $^1J_{NH}$ values from random coil values, being smaller than −0.35 Hz		[201]
$^1J_{NC'}$	(–)13–15	Correlation with backbone amide H-bonds in different secondary structures		[203]
$^1J_{C\alpha C\beta}$	(+)32–38	Correlation with ϕ and ψ backbone dihedral angles: $\Delta^1J_{C\alpha C\beta} = 1.3 + 0.6\cos(\psi - 61°) + 2.2\cos[2(\psi - 61°)] - 0.9\cos[2(\varphi + 20°)] \pm 0.5$ Hz where $\Delta^1J_{C\alpha C\beta}$ is the deviation from $^1J_{C\alpha C\beta}$ in the corresponding free amino acid		[204]
$^1J_{C\alpha H\alpha}$	(+)130–150	Correlation with ϕ and ψ backbone dihedral angles: $^1J_{C\alpha H\alpha} = A + B\cos^2(\phi - 30°) + C\cos^2(\psi - 30°)$ where A, B and C are empirically-derived coefficients		[205]
$^2J_{H\alpha H\beta}$	(+)15–19	Torsional ψ angle in glycines		[200]
$^3J_{HN-H\alpha}$	(+)0–11	Backbone ϕ angle	$^3J = A\cos^2\theta - B\cos\theta + C\sin^2\theta$ (The Karplus relationship)	[200,206]
$^3J_{H\alpha H\beta}$	(+)1–12	Side-chain χ_1 angle		
$^{3h}J_{NC'}$	(–)0.5	Cross H-bond coupling: $^{3h}J_{NC'} = (-357\text{ Hz})\exp(-3.2r_{HO})\cos^2\theta$ where r_{HO} represents the H···O distance (Å) and θ the H···O=C angle		[207]

Table 1.1: Nuclear spin-spin couplings in proteins

Δ^1J_{NH} measurement has also enabled the detection of a transient H-bond in the folding nucleus of the coiled-coil GCN4 leucine zipper [208]. GCN(16–31) comprises the trigger sequence, a short autonomous helical folding unit which is indispensable for coiled-coil formation. Arg-25 in native GCN(16–31) has the most negative Δ^1J_{NH} value of –0.46 ± 0.09 Hz. Given the small experimental error of ~0.05 Hz estimated by recording the NMR experiment at least twice for each sample, and assuming a rigid helix has Δ^1J_{NH} of –1.0 Hz,

Arg-25 is suggested to be involved in an intramolecular H-bond for ~46% of the time. This transient H-bond is validated by H/D exchange [208], and is consistent with the results from NMR and CD studies [209]. Δ^1J_{NH} values at other residues of GCN(16–31) correlate well with trifluoroethanol-induced helical structures, and show the lack of helical structures in the presence of GdmCl and in the R25A mutant, in which Arg-25-mediated electrostatic interactions that are essential for GCN stability are abolished [208]. Overall, $^1J_{NH}$ measurements have been demonstrated to be highly sensitive tools for the detection of stable and transient H-bonds in small soluble proteins.

1.4 Aims

Rhodopsins are photochemically reactive membrane proteins that covalently bind to retinal chromophores. Type I rhodopsins are found in prokaryotes and eukaryotic microbes, and function as light-driven ion transporters (e.g. bacteriorhodopsin for pumping H^+ and halorhodopsin for pumping Cl^-), or phototaxis receptors (e.g. sensory rhodopsins I and II). Type II rhodopsins consist of photosensitive G protein-coupled receptors (GPCRs) in animal vision (e.g. bovine rhodopsin), and also in the pineal gland, hypothalamus and other tissues of lower vertebrates [210,211]. All proteins in the rhodopsin family harbour a seven transmembrane (7TM) α-helical fold and a Schiff base linkage between the ε-amino group of a conserved lysine residue on helix G (or Helix 7) and the retinal chromophore. Despite structural similarities between bacteriorhodopsin and bovine rhodopsin, there are noted mechanistic differences between the two proteins. The retinal chromophore in archaeal rhodopsins photoisomerises and remains covalently attached to the protein throughout the photocycle before the chromophore re-isomerises and the protein returns to the ground state [212]. On the other hand, retinal in bovine rhodopsin photoisomerises from 11-*cis* to all-*trans*, dissociates from the protein, and is regenerated via a series of enzymatic steps to 11-*cis* retinal, which rebinds to opsin to regenerate rhodopsin [213].

The unfolding/folding pathways between bacteriorhodopsin and bovine rhodopsin are highly different, leading to different folding models being proposed. One of the major unanswered questions in the field is how the same heptahelical transmembrane architecture gives rise to different responses to denaturants and the proposal of different folding models. Most α-helical membrane proteins which have been studied to date appear to adhere the two-stage model. Studies with new model proteins and across comparable experimental conditions are required to further explore the long-range interactions model and the three-state model for α-helical membrane protein folding.

The aims of this thesis broadly involve the investigation of the folding mechanism of a 7TM α-helical membrane protein, sensory rhodopsin II, and further development and application of methods for membrane protein folding studies. Specific aspects of these broad aims are addressed in separate chapters of this thesis:

1. To screen for suitable *in vitro* conditions and to characterise the unfolding and refolding kinetics of pSRII using biophysical techniques (Chapter 2);
2. To characterise the denatured states of pSRII using biophysical techniques and solution-state NMR (Chapters 2 & 3);
3. To investigate the application of $^{1}J_{NH}$ scalar coupling measurements for detecting H-bonds in membrane proteins (Chapter 4);
4. To further explore the possibility of assembling membrane proteins with split inteins using pSRII as proof-of-concept (Chapter 5).

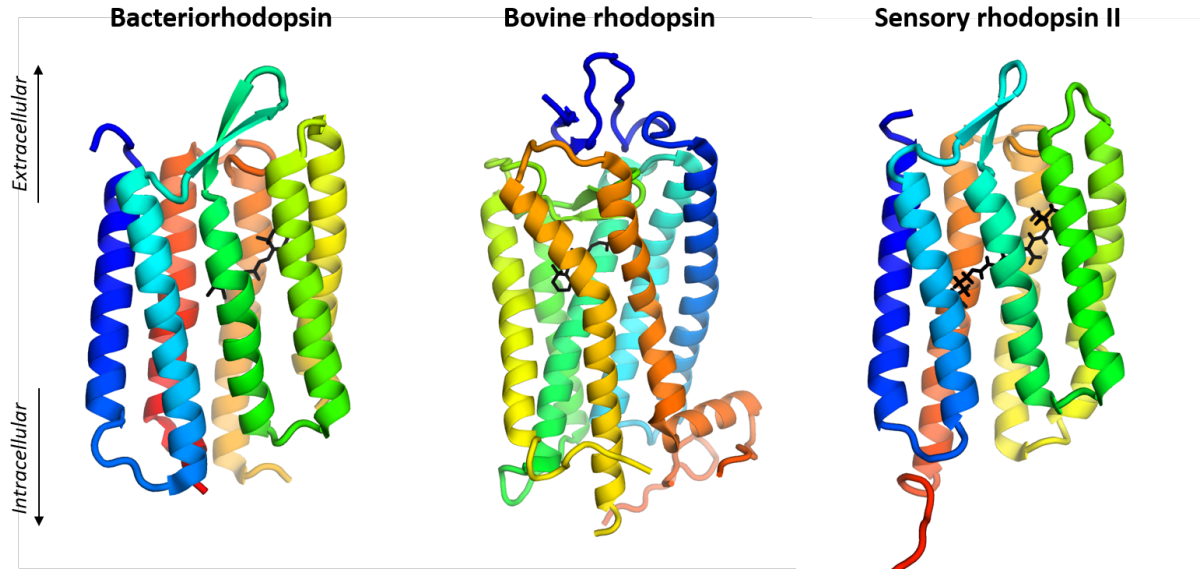


Figure 1.8: 3D structures of bacteriorhodopsin, bovine rhodopsin and sensory rhodopsin II
1.55 Å crystal structure of *H. salinarium* bacteriorhodopsin (PDB 1C3W) [214]; 2.6 Å crystal structure of bovine rhodopsin (PDB 1L9H) [215]; and solution-state NMR structure of *N. phararonis* sensory rhodopsin II (PDB 2KSY) [216]. The retinal chromophore in each protein is shown as black sticks.

Sensory rhodopsin II (pSRII) (Figure 1.8), from the archaeon *Natronomonas pharaonis*, functions as a repellent phototactic receptor to blue light. The photocycle of pSRII begins with photoisomerisation of the all-*trans* retinal chromophore to the 13-*cis* conformation, proceeding through a series of photointermediates (K, L, M, N and O) and initiating a series of conformational changes in the seven transmembrane helix core [217]. Proton transfer from the protonated Schiff base to its counterion, Asp-75, occurs upon the formation of the M-state, which is the signalling state [218,219]. The decay of the O-state coincides with proton transfer from the protonated Asp-75 to the proton releasing group, Asp-193, located at the

extracellular surface through an extracellular proton conduction channel (consisting of Pro-182, Pro-183 and Val-194 in pSRII) [220]. Proton uptake occurs subsequently, leading to reprotonation of the Schiff base and enabling pSRII to exhibit weak outward proton pumping activity in the absence of pHtrII (the cognate transducer of pSRII) [221,222]. The photocycles and photointermediates of different archaeal rhodopsins are tuned to the function of the protein: in the case of pumps such as bacteriorhodopsin, the photocycle time is fast (several milliseconds) [223], consistent with rapid proton pumping, whereas in sensory rhodopsins such as pSRII, the photocycle lasts longer (~ 1 s) [224], allowing signal transfer.

The two transmembrane helices (TM1 and TM2) of pHtrII interact with helices F and G of pSRII to form a 2:2 complex with the two transducer molecules at the center of the complex. Light activation of pSRII triggers an outward movement of the cytoplasmic side of helix F, leading to a rotation of TM2 and the unwinding of helices in the cytoplasmic region of pHtrII [225]. This conformational change is recognised by the homodimeric histidine kinase CheA via a coupling protein CheW. CheA undergoes autophosphorylation and activation, and phosphorylates CheY, an aspartate kinase. Phosphorylated CheY interacts with the flagellar motor to induce clockwise rotation of the flagellum, enabling the archaeon to tumble away from light. Photoadaptation (sensitivity of the response to light activation) is regulated by CheB (a methylesterase) and CheR (a methyltransferase): autophosphorylated CheA also phosphorylates CheB, which competes against CheR to control the methylation state of specific glutamate residues in the cytoplasmic coiled-coil domain of pHtrII [226]. Ssignalling continues until the O-state photointermediate decays back to the ground-state conformation.

Simulated thermal unfolding experiments suggest that pSRII is stabilised by a combination of mechanisms found for rhodopsin and bacteriorhodopsin [139], with a subset of helices breaking off early in the simulation and a core involving residues from other helices and loops remaining until late. More experimental studies are needed to understand the folding pathway of membrane proteins, particularly how and when secondary and tertiary structure elements are being formed. Retinal-binding 7TM helical proteins are excellent models for these studies, since the retinal chromophore acts as a natural reporter of the binding pocket and thus, the tertiary structure.

To address our knowledge gap on how different membrane proteins traverse their folding landscapes, the SDS-denatured states and kinetics for the unfolding and refolding of pSRII in SDS/diheptanoylphosphatidylcholine (c7-DHPC) mixed detergent micelles were investigated in Chapter 2. Extensive screening of different denaturants and additives identified SDS as the most suitable denaturant for studying the unfolding of pSRII. Circular dichroism, UV/visible spectroscopy and tryptophan fluorescence showed that the SDS-denatured state is characterised by loss of tertiary structure (inter-helical contacts) and little loss in secondary structure, similar to bacteriorhodopsin and hence distinguishing pSRII from the long-range interactions model. The native chromophore can be reformed from a SDS-denatured state with retinal remaining covalently attached and from the apo-protein state, showing that both protein folding and retinal attachment are reversible. Notable differences were found in the kinetics of SDS denaturation and the refolding of pSRII compared to bacteriorhodopsin and other membrane proteins, which fold with two-state kinetics. These results suggest marked differences in the unfolding and folding pathways amongst pSRII, bacteriorhodopsin and bovine rhodopsin. Overall, this chapter has established pSRII as a new model protein for membrane protein folding studies.

Solution-state NMR offers distinct advantage in providing structure and dynamics information at atomic resolution for highly-populated ground states as well as transient minor populations in bio-macromolecules. A range of NMR methods are available for studying protein folding events which occur at different timescales and for probing the properties of low-population folding intermediates, hence yielding detailed insights into folding landscapes, misfolding, aggregation and function [227]. In Chapter 3, SDS-denatured pSRII, acid-denatured pSRII and apo-sensory opsin obtained by hydroxylamine-mediated bleaching of pSRII were characterised by solution-state NMR. Small changes in the chemical environment of backbone amides were detected, consistent with the view that denatured pSRII remains embedded in detergent micelles. Changes in backbone amide dynamics, specifically increased observation of intermediate timescale exchange, was an important feature of SDS- and acid-denatured pSRII and apo-sensory opsin. This chapter has shed insight on differences in unfolding pathways induced by SDS and acidic pH, and the crucial roles of the retinal chromophore in governing the structural integrity and dynamics of the pSRII helical bundle.

Different methods for measuring $^1J_{NH}$ scalar coupling were explored in Chapter 4 with the vision of applying these methods for detecting H-bond formation during membrane protein folding. During protein folding, the possible presence of stable and transient H-bonds might contribute differently towards stabilising native-like or non-native structures in intermediate states. Existing studies in literature (see Section 1.3.2) have been limited to small soluble proteins of < 10 kDa. RalA, a 20.9 kDa small GTPase with a mixed α + β fold, was used in proof-of-concept studies for extending the molecular weight limit and assessing the precision and accuracy of different scalar coupling measurement methods. The applications of $^1J_{NH}$ scalar coupling measurements for studying hydrogen bonds in folded and denatured membrane proteins were then investigated using OmpX reconstituted in DPC micelles and in the urea-denatured state. The applicability of the explored methods for measuring $^1J_{NH}$ in detergent-solubilised membrane proteins was limited by large errors attributable to excessive cross-correlated relaxation. Further optimisation of NMR experiments and utilisation of deuterated samples will be required in future studies.

The two-stage model and the long-range interactions model for α-helical membrane protein folding are both supported by studies on protein assembly using different complementary fragments. Assembling isolated protein fragments requires solvent conditions which are suitable for solubilising all fragments and maintaining functionality of the native state. Naturally-occurring split inteins can mediate spontaneous *trans*-splicing of protein fragments both *in vivo* and *in vitro*, hence offering discernible advantages over protein assembly from isolated fragments in terms of versatility of the method and potential comparability to folding *in vivo*. Successful assembly of correctly-folded and functional proteorhodopsin has been achieved by *in vivo trans*-splicing of two complementary fragments which are co-expressed in *E. coli* [228]. With the vision of applying this method toward understanding the roles of helices and loops in the folding of pSRII and other membrane proteins, in Chapter 5, pSRII assembly *in vivo* was attempted by introducing a splice site in the loop region of the β-hairpin constituting the BC loop of pSRII. Despite attempts at optimising the conditions for protein expression, further optimisation is required for successful *in vivo trans*-splicing of pSRII to yield the native chromophore.

2 Biophysical characterisation of denatured states and reversible unfolding of sensory rhodopsin II

2.1 Introduction

Protein denaturation *in vitro* offers an avenue for accessing conformations represented in the ensemble of unfolded states. While early studies of proteins unfolded with high concentrations of chemical denaturants or at high temperatures initially led to the impression that unfolded states are essentially random coils, subsequent detailed studies on denatured states under a variety of solvent conditions have highlighted the influence of both native and non-native clustering of aliphatic and aromatic side-chains on the folding landscape [229]. Understanding the residual structure in the chemically denatured state of membrane proteins might provide mechanistic clues to how they fold. While different folding models are accepted for soluble proteins, only the two-stage model [86] and the long-range interactions model [140] have been proposed so far for helical membrane proteins. Simulated thermal unfolding experiments suggest that pSRII is stabilised by some combination of native-like secondary structure in a subset of helices and a folding core involving tertiary structure interactions between other helices and loops [139].

To address the knowledge gap on how different membrane proteins traverse their folding landscapes, this chapter focuses on biophysical characterisation of the SDS-denatured states and the kinetics for unfolding and refolding pSRII in SDS/c7-DHPC mixed detergent micelles, with the vision of proposing unfolding and folding mechanisms for pSRII. Experimental results on pSRII are compared against those in literature for bacteriorhodopsin and bovine rhodopsin, with the aim of comparing the unfolding and folding pathways of the three retinal-binding proteins.

Characterisation of the denatured states and the unfolding and folding pathways require maximal unfolding of the protein with minimal amount of aggregation. However, these conditions are difficult to fulfil for membrane proteins, since *in vitro* studies require

membrane proteins to be functionally reconstituted in a membrane mimetic, which often also renders them resistant to denaturation. It is particularly challenging to minimise the aggregation of denatured membrane proteins given their high composition of hydrophobic residues and the exposure of hydrophobic regions during denaturation. Moreover, experimental conditions that have been successfully applied to unfold a protein might not be applicable to others; hence, no uniformly-suited approach has been established. This chapter therefore begins with extensive screening of denaturants and additives to identify the most suitable solvent conditions for studying the unfolding of pSRII (see Section 2.3.1).

To understand the unfolding mechanism, various biophysical techniques, including circular dichroism, UV/visible spectroscopy and tryptophan fluorescence were used to characterise changes in the secondary and tertiary structures and the timescales at which these changes occur upon SDS-denaturation of pSRII (see Sections 2.3.2, 2.3.3 and 2.3.6).

The refolding pathway was examined in Sections 2.3.4–2.3.5. The reversibility of SDS denaturation and refolding kinetics were studied by monitoring changes in the absorbance of the retinal chromophore using UV/vis spectroscopy. Refolding was initiated from SDS-denatured pSRII with covalently attached retinal chromophore and from the apo-protein state. These studies would shed insight on the reversibility of protein folding and retinal Schiff base hydrolysis, respectively.

Overall, this study has established pSRII as a new model protein for membrane protein folding studies, and has shown that pSRII possibly has a unique folding mechanism which is distinct from those of bacteriorhodopsin and bovine rhodopsin.

2.2 Theory

2.2.1 Circular dichroism

Circular dichroism (CD) spectroscopy is a well-established tool for the structural characterisation of different biopolymers, including proteins, carbohydrates and nucleic acids. The physical basis for CD is the differential absorption of left- and right-handed circularly polarised light (Figure 2.1) by asymmetric (chiral) molecules with a chiral centre or a 3D structure that provides a chiral environment. The difference in absorption (ΔA) of left- (A_L) and right-handed (A_R) circularly polarised light is measured as a function of wavelength to yield the CD spectrum of the sample:

$$\Delta A = A_L - A_R \tag{2.1}$$

CD data is commonly reported in terms of mean residue ellipticity (MRE or [θ]; degree cm^2 dmol^{-1} residue^{-1}) or delta epsilon ($\Delta\varepsilon$; M^{-1} cm^{-1}). Ellipticity (the machine unit) measures the difference in molar extinction coefficients between left- and right-handed circularly polarised light, and needs to be corrected for the amount of protein in the sample.

$\Delta\varepsilon$ is the difference in molar extinction coefficients:

$$\Delta\varepsilon = \frac{\theta}{10 \times l \times c \times n \times 3298} \tag{2.2}$$

where θ is the observed ellipticity (millidegrees), l the path length (cm), c the protein concentration (M) and n the number of peptide bonds.

MRE ([θ]) is calculated as:

$$[\theta] = 3298 \times \Delta\varepsilon \tag{2.3}$$

Information on protein secondary structure can be derived from CD signals in the far UV (190–240 nm) wavelength range due to the amide chromophores of the peptide bonds. The two types of electron transitions responsible for the CD signals in this wavelength region are n→π* transitions at around 222 nm and π→π* transitions at around 208 nm and 190 nm.

Figure 2.1b shows the far-UV CD spectra of the secondary structures which are commonly found in proteins and peptides. α-helices are characterised by two minima at 222 nm and 208 nm, and β-strands are characterised by a single minimum at 218 nm [230].

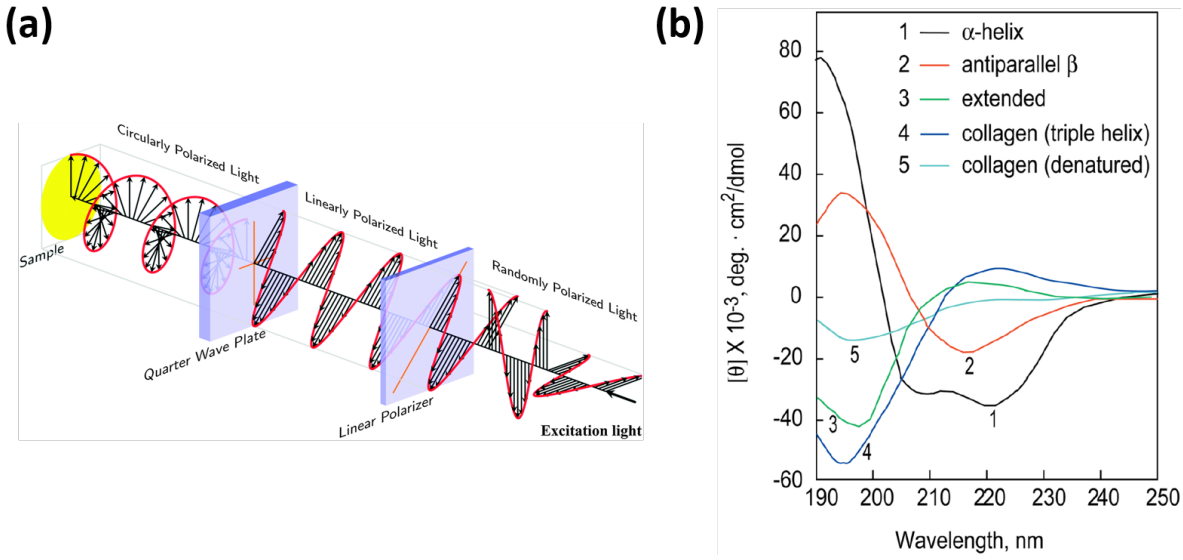


Figure 2.1: Circular dichroism
(a) Schematic diagram showing the differences between unpolarised (randomly polarised) light, linearly polarised light and circularly polarised light. Unpolarised light contains oscillations of the electronic component in all directions perpendicular to the direction of propagation. The direction of the electronic component is restricted to a plane perpendicular to the direction of propagation in linearly polarised light. In circularly polarised light, the magnitude of oscillation is constant and the direction oscillates in a clockwise or anti-clockwise manner. Retrieved from [231]. **(b)** CD spectra of polypeptides and proteins with representative secondary structures: poly-L-lysine in the α-helical (1) and antiparallel β-sheet conformations (2) at pH 11.1, and in the extended conformation (3) at pH 5.7; placental collagen in its native triple-helical (4) and denatured (5) forms. Retrieved from [230].

Far-UV CD is commonly used to assess protein secondary structural changes across various applications, such as ligand binding, mutations, protein denaturation and refolding by chemical and thermal methods. Changes in ellipticity at the characteristic wavelengths of different secondary structure elements can be monitored as a function of temperature, concentration of denaturants or stabilising agents, ligands or time.

Measurements at single wavelengths are useful for following the kinetics and thermodynamics of the folding of polypeptides and proteins, for example, calculating helical content from ellipticity values at 222 nm and 208 nm. Monitoring changes at a single

wavelength enables data to be collected rapidly, although the information content is limited and other conformations may interfere with the estimation of α-helical content [232].

Information on tertiary structure can be obtained from CD spectra of protein aromatic residues in the near UV (250–300 nm) wavelength range. CD bands in the near UV and visible region (> 300 nm) are also attributed to extrinsic chromophores. While the extrinsic chromophores may not be chiral or may exist as enantiomeric mixtures, their interactions with the chiral environment of a protein binding site generates optical activity [233]. Analyses of CD bands using a combination of computational and experimental methods can therefore yield information about the conformation of the chromophore and the binding pocket. An extensively studied example is the Soret CD band in the 400-nm region arising from the porphyrin ring in heme proteins, where the CD band can be explained by the combined effects of porphyrin-protein coupling and structural distortions of porphyrin [234]. Computational studies on visual rhodopsins show that nonplanar distortions in the polyene chain of retinal protonated Schiff base (PSB) are dominant over coupling of the retinal-PSB to the peptide backbone and aromatic side chains of the protein [235].

Whereas soluble proteins form uniform isotropic solutions in aqueous environments, membrane proteins are usually reconstituted into membrane mimetics to yield large hydrophobic, anisotropic particles. The physical nature of these samples might give rise to spectral artifacts which could present problems during data collection, analysis and interpretation [232]. Membrane, lipid and detergent environments tend to be hydrophobic and/or amphipathic. Hence, membrane protein samples have different physical properties and produce different spectral characteristics compared to soluble proteins. Secondary structure analyses usually entail empirical methods based on reference datasets of spectra derived from proteins of known structures. Peak positions in CD spectra can be affected by solvent dielectric. This is relevant for membrane proteins, due to the considerably lower dielectric constant of the hydrophobic core in a detergent micelle, lipid or membrane bilayer environment compared to that of aqueous solution. The extent and nature of this shift depends on the electronic transition in the peptide and on the relative position of the peptide bond inside the membrane environment. The wavelength dependence on solvent dielectric is thus non-linear. Shifts in peak positions can have significant effects on secondary structure

analyses, and tend to produce inaccurate results with reference datasets derived from soluble proteins for spectra deconvolution. An effective solution is to use a reference dataset containing the spectra of membrane proteins, for example the SMP56 reference set from CDPro [236] and the SMP180 reference dataset [237] available on the DichroWeb server.

The software CDPro was used in this chapter to analyse secondary structure composition of pSRII in different solvent conditions. All methods of deconvoluting CD spectra assume that the spectrum can be represented by a linear combination of its secondary structural elements plus a noise term. Protein conformation is generally evaluated against the spectra of references – a set of proteins with known conformations determined by X-ray crystallography [238]. CDPro includes 3 different programs for modelling CD spectra of unknown samples against the references: CDSSTR, CONTIN and SELCON3. The versions of these programs provided in CDPro have been modified to accept any given set of reference proteins, including the provided reference sets. In the variable selection method implemented in CDSSTR, protein spectra are eliminated systematically from the set of standards to create reduced data sets – effectively, a reduced set is created to include only reference proteins that fulfill the selection criteria for a good fit. Proteins in the reference set can also be weighed variably in a least-squares fitting of the analysed CD spectrum, as is done in the ridge regression method implemented in CONTIN [236].

2.2.2 UV/visible spectroscopy

UV/visible (UV/vis) absorbance spectroscopy is routinely used in analytical chemistry for quantitative analyses of different samples, including biological macromolecules. Molecules containing π-electrons or non-bonding electrons can absorb energy in the form of ultraviolet or visible light to excite these electrons to higher anti-bonding molecular orbitals via $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ transitions, respectively. An absorption spectrum thus measures the transition of these electrons from the ground state to the excited state. A chromophore is the part of a molecule which absorbs light at a particular frequency to impart colour to the molecule. The wavelengths of absorption peaks can be correlated with the chemical structure. The energy gap and maximum absorbing wavelength (λ_{max}) associated with the $\pi \rightarrow \pi^*$ transition vary with the substituents on the double bonds and the length of conjugation. For example,

lycopene and β-carotene have extended conjugated systems of 11 double bonds and absorb at ~500 nm [239].

Free all-*trans* retinal absorbs at 381 nm in ethanol [240] due to the conjugated system formed by its polyene chain. Retinal is bound to the ϵ-amino group of lysine via a protonated Schiff base (PSB) linkage. PSB of both the *cis*- and *trans*-forms of retinal absorb at 440 nm in methanol [241]. The "opsin shift" is defined as the difference in the absorption maxima between free retinal and the corresponding protonated Schiff base model compound in methanol [242]. The value of the opsin shift depends on the conformational differences between protein-bound retinal and free retinal in solution, as well as contributions of the protein binding pocket towards the absorption properties of the chromophore. Hence, the sensitivity of the UV/vis absorption profile of retinal to the local environment can be used as a probe for specific protein-chromophore interactions.

Figure 2.2 shows the absorption spectra of different retinal isomers bound to visual and bacterial rhodopsins. All-*trans* retinal bound to pSRII absorbs at 498 nm, along with two vibronic side bands at 460 and 420 nm, and a third vibronic side band at 370 nm which is derived from multi-Gaussian deconvolution of the spectrum [224]. Gaussian band-shape simulation with vibrational frequencies of polyene stretching modes reproduces the structured absorption spectra of native pSRII pigment and the analogue 3,4-dehydroretinal [243]. The absorption spectra of 11-*cis* retinal in bovine rhodopsin (λ_{max} = 498 nm) [244] and all-*trans* retinal in light-adapted bacteriorhodopsin (λ_{max} = 568 nm) [245] appear as simple bell shapes with no vibrational fine structures.

Studies on absorption shifts produced by different retinal analogues indicate that planarisation of the retinal ring with respect to the polyene chain is the dominant factor in wavelength tuning of retinal pigments in the rhodopsin protein family. The small opsin shift in pSRII compared to those of other bacterial rhodopsins is due to both the lack of electrostatic perturbation of the chromophore and the relatively strong proton-counterion interaction at the Schiff base nitrogen atom [243].

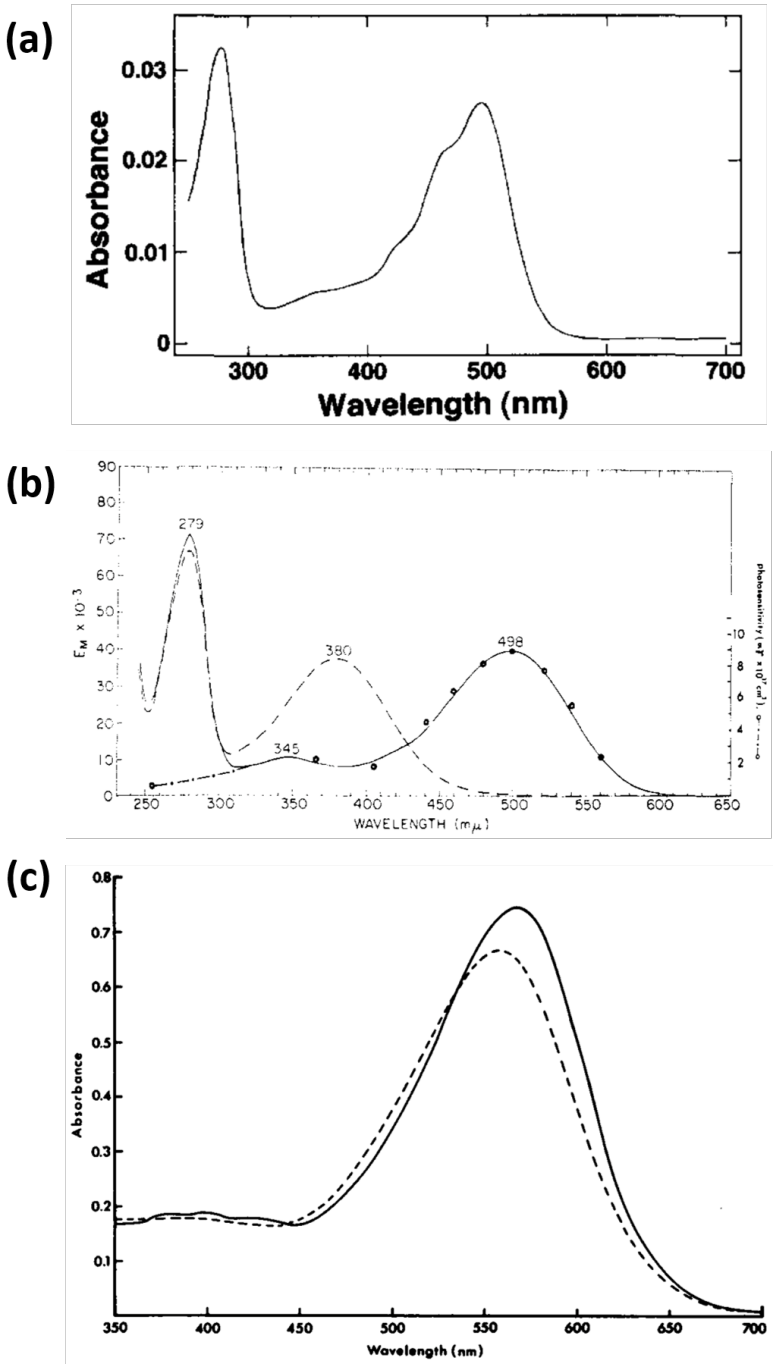


Figure 2.2: Absorbance spectra of pSRII, bovine rhodopsin and bacteriorhodopsin
(a) Purified pSRII in 0.28% octyl glucoside, 2 M NaCl, 25 mM PIPES, pH 7.2 [246]. **(b)** Bovine rhodopsin in 0.04 M CTAB, 0.066 M potassium phosphate, pH 6.5: (—) unbleached; (---) bleached; (-•-) photosensitivity curve [247]. **(c)** Purple membrane of *Halobacterium halobium*: (—) light-adapted membrane; (---) dark-adapted membrane [248].

2.2.3 Spectral deconvolution

Resolving a complex spectrum into individual bands is useful for yielding information about each band. Spectral deconvolution involves fitting of a nonlinear model to the input data followed by iterative refinement. With each iteration the input is approximated by a linear combination of its components. The least squares method is used in regression analysis to approximate the solution of overdetermined systems (sets of equations in which there are more equations than unknowns). The solution minimises the sum of the squares of the residuals.

The Levenberg-Marquardt algorithm [249,250] is an iterative technique which locates the minimum of a multivariate function expressed as the sum of squares of non-linear real-valued functions. The algorithm can be thought of as a combination of two minimisation methods: the steepest descent method and the Gauss-Newton method. In the steepest descent method, the local minimum is located by reducing the sum of the squared errors whilst taking steps proportional to the negative of the gradient at the current point. The steepest descent method is relatively slow near the local minimum. The Gauss-Newton method reduces the sum of the squared errors by assuming the least squares function is locally quadratic, and searches for the minimum point of the quadratic. The Levenberg-Marquardt method behaves more like the steepest descent method when the starting parameters are far from their optimal values and more like the Gauss-Newton method when the parameters are close to their optimal values [251].

A symmetrical spectral band is described by three parameters: position (maximum absorbing wavelength or λ_{max}), intensity, and width (usually the bandwidth at half-height). The band shapes most commonly used for deconvolution are the Gaussian function and the Lorentzian function, both of which are symmetrical functions. UV/vis absorbance spectra generally have Gaussian band shapes, while the Lorentzian function is usually used for simulating NMR spectra [252]. The result of a multi-gaussian fit of the absorbance spectrum of pSRII is shown in Figure 2.3 [224].

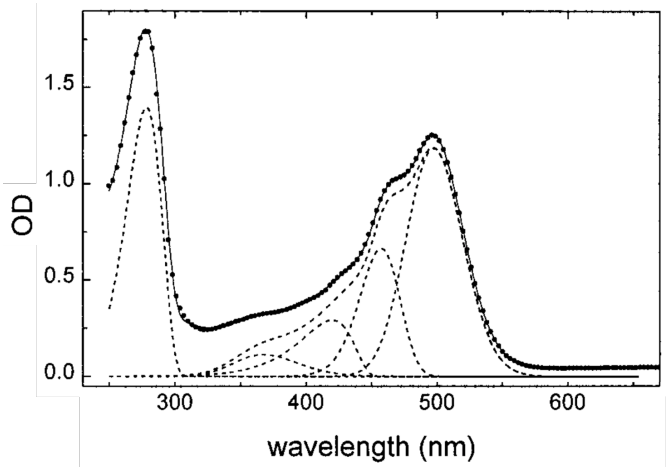


Figure 2.3: Multi-Gaussian fit of the absorbance spectrum of purified pSRII
Spectrum of purified pSRII reconstituted in *n*-octyl-β-D-glucopyranoside (OG). The spectra of individual fitted components are shown as dashed lines, and the sum of all components are shown as solid line. Adapted from [224].

2.2.4 Tryptophan fluorescence

Fluorescence is the re-emission of light leading to the return of a molecule to a lower energy level after the absorption of electromagnetic radiation (e.g. visible light). In terms of the Jablonski diagram (Figure 2.4a), the initial absorption of a photon by a fluorophore leads to promotion from the ground vibrational state of S_0 to various excited vibrational states of S_1. Such transitions are said to be vertical, in the language of the Franck-Condon principle. Non-radiative relaxation to the ground vibrational state of S_1 then takes place. Fluorescence arises when a photon is emitted during the transition from the ground vibrational state of S_1 down to various vibrational states of S_0. Since energy is lost through non-radiative mechanisms before reaching the emitting state, the fluorescence spectrum is shifted to higher wavelengths (lower energies) compared to the absorption spectrum. The quantum yield (η), a measure of how efficiently absorbed photons are converted into emitted light, determines how "fluorescent" a fluorophore is:

$$\eta = \frac{\text{number of emitted photons}}{\text{number of absorbed photons}} \tag{2.4}$$

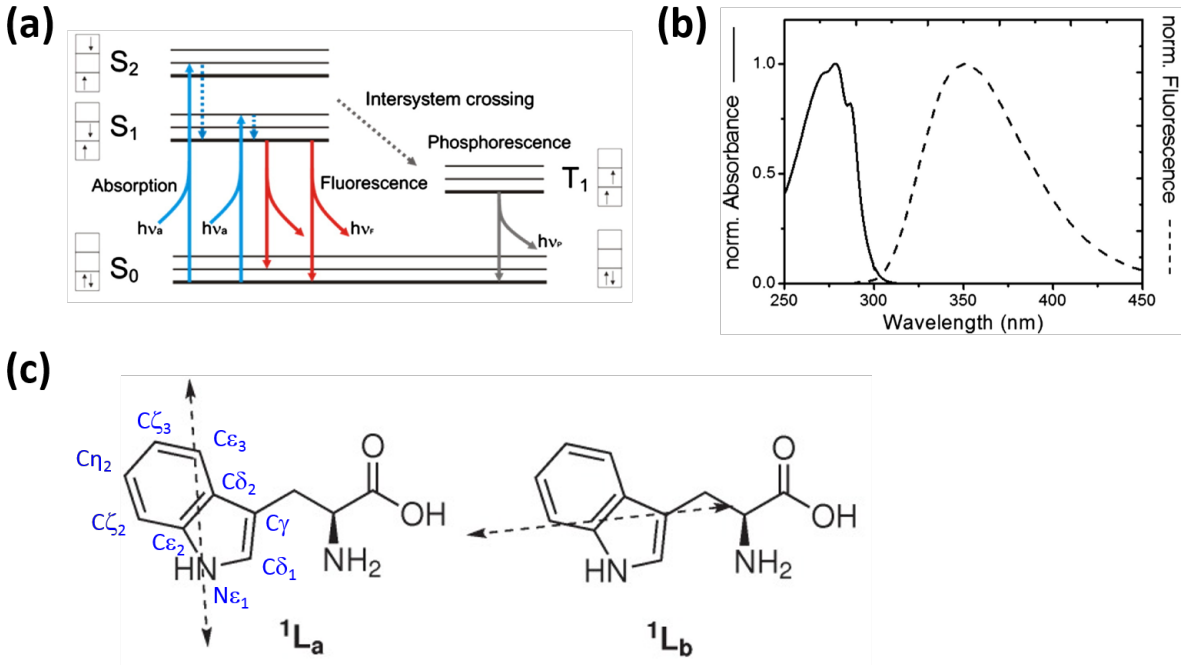


Figure 2.4: Tryptophan fluorescence
(a) Jablonski diagram illustrating possible transitions between electronic states of a fluorophore during absorption (solid blue) and emission (solid red) of a photon. Non-radiative relaxation is shown in dashed blue, and intersystem crossing/phosphorescence in gray. **(b)** Absorption and emission spectra of tryptophan (excitation wavelength = 280 nm). Retrieved from [253]. **(c)** Electronic absorption transitions of tryptophan, ^1L_a and ^1L_b. The nomenclature for atoms in the tryptophan indole ring are labelled in blue and the directions of the dipole moments in the ^1L_a and ^1L_b state are shown by the dotted arrows. Adapted from [254].

The fluorescence emission spectrum is usually recorded by exciting the sample at a fixed wavelength and then measuring the emission at different wavelengths. The emitted light is detected perpendicular to the direction of the exciting beam to avoid swamping the detector with very intense radiation normally used for excitation. Similarly, the fluorescence excitation spectrum can be recorded by scanning the wavelengths of exciting radiation while detecting at a fixed wavelength.

Fluorophores can be available intrinsically through the presence of fluorescent amino acids (phenylalanine, tryptophan and tyrosine) or chromophores, or extrinsically by covalent or non-covalent attachment of fluorescent dyes or fluorescent proteins. Proteins from the rhodopsin family have two intrinsic fluorophores – the covalently-bound retinal chromophore and tryptophan residues. Bound retinal is weakly fluorescent due to its low quantum yield on

the order of 10^{-4} to 10^{-5}, originating from the fast return to the ground state in the ps–fs timescale after retinal excitation [255].

Tryptophan is the most abundant (up to 3% in membrane proteins) amongst the three fluorescence amino acid constituents [254]. Most retinal proteins contain several tryptophan residues, for example bovine rhodopsin has five and pSRII has six tryptophan residues. The indole ring of tryptophan has two absorption/emission transitions, ^1L_a and ^1L_b (Figure 2.4c). The current consensus is that tryptophan fluorescence emission originates from only the ^1L_a state, except when the local environment is completely nonpolar [254]. Upon excitation of tryptophan, there is large redistribution of electron density from the $N\varepsilon_1$ and $C\gamma$ atoms to the $C\varepsilon_3$, $C\zeta_2$, and $C\delta_2$ atoms of the asymmetric indole ring, causing a large dipole moment in the ^1L_a excited state. The photophysical properties of tryptophan are influenced by the polarity of its local environment, hydrogen bonding and other non-covalent interactions. This is primarily attributed to interaction of the induced dipole moment in tryptophans with dipoles in the local environment influenced by solvent or protein side-chains. Computational studies suggest that the relative orientation of the charge or dipole to the ring determines whether there will be a red shift: positively charged moieties near the benzene end or negative charged moieties near the pyrrole end of the tryptophan ring system will produce a red shift (longer wavelengths) [256]. The emission λ_{max} is blue-shifted for tryptophans in non-polar environments, for example Trp-45 of azurin is buried deep into a hydrophobic pocket and has λ_{max} of 308 nm [257]. In contrast, solvent-exposed tryptophans have the most red-shifted emission, for example, Trp-25 in glucagon monomer emits with λ_{max} of 352 nm [258].

In comparison, other fluorescent amino acids have smaller contributions towards the intrinsic fluorescence emission of a protein. The contribution of phenylalanine is negligible due to its very low quantum yield. Tyrosine has similar quantum yield and absorption wavelength as tryptophan [259]. Illumination at 280 nm excites both tyrosine and tryptophan residues. Selective excitation of mostly tryptophan residues can be achieved by using excitation wavelengths within the range of 295–305 nm. This is because the absorption properties of tryptophan are governed by large redistribution of electron density in the asymmetric ring system upon excitation, whereas little redistribution of electron density occurs in the symmetric ring systems of tyrosine and phenylalanine [260,261].

Tryptophan fluorescence in retinal proteins is quenched due to efficient energy transfer to the retinal by Förster resonance energy transfer (FRET), as there is considerable overlap between tryptophan fluorescence emission and the absorbance spectrum of retinal in different states (free, PSB and protein-bound native pigment) [107]. Studies on bacteriorhodopsin excited at 280 nm indicate that energy transfer from excited tyrosines and tryptophans to the chromophore takes place with a quantum yield of 0.7–0.8. All tyrosine residues and most tryptophan residues are completely quenched by the transfer process, with one tryptophan remaining unquenched and another being about 80% quenched [262].

The high dependency of tryptophan fluorescence on its local environment has been utilised extensively to study protein conformational changes and interactions with binding partners. Unfolding and refolding studies monitored by tryptophan fluorescence were performed for bacteriorhodopsin and bovine rhodopsin to study membrane protein folding. Information about the residual structure in SDS-denatured states of bovine rhodopsin suggest two different stages of unfolded structures at low and high concentrations of SDS, respectively [135]. Tryptophan fluorescence kinetics studies on the refolding of bacteriorhodopsin reveal discrete steps for retinal binding to an apo-protein intermediate and for subsequent formation of the covalent Schiff base linkage [120].

Although tryptophan fluorescence is highly utilised to study protein conformational changes, the interpretation of tryptophan fluorescence spectra can be challenging since different factors are coupled to yield the observed spectral changes. Moreover, most proteins possess multiple tryptophan residues, and the overall emission yields only averaged information on the protein structure. Extracting the contribution of each tryptophan reporter to obtain structural information on different parts of the protein is difficult.

Correlation of the observed changes in tryptophan fluorescence with protein structural changes is not necessarily straightforward. A study on a fragment of bacteriorhodopsin containing helices A and B using tryptophan fluorescence and molecular dynamics simulation shows that the observed decrease in FRET efficiency between Tyr-57 on helix B and Trp-10 and Trp-12 on helix A is due to a decrease in the quantum yield of tryptophans,

attributed to a cluster of water molecules forming a hydrogen-bond network with Tyr-57, rather than disruption of interhelical interactions [263]. This study indicates that changes in tryptophan fluorescence does not necessarily indicate structural changes, but may also be explained by changes in the photophysics of the fluorophores.

Another major concern when studying retinal proteins is that probing the fluorophore with excitation light could cause photoactivation of the retinal. The exact effect depends on the nature of the photocycle. Visual rhodopsin is bleached even under continuous illumination of low (sub-μW/cm^2) light intensity for up to 1 h [264], causing the chromophore to convert from the dark-state (λ_{max} = 498 nm) to the active meta-II state (λ_{max} = 380 nm). The light sensitivity of visual rhodopsin relates to its function of enabling vision in low-light conditions [253]. Bacteriorhodopsin photobleaches after continuous illumination for at least 30 min by yellow light with an intensity of 100–400 mW/cm^2, although there is little bleaching at up to 30 °C at pH 7 [265]. Continuous illumination of pSRII reconstituted in *n*-octyl-β-D-glucopyranoside (OG) leads to a stable M-like state which binds 13-*cis* retinal and absorbs at 386 nm. Overall, it is therefore recommended to use very tight excitation slits to achieve low excitation and wide emission slits of > 10 nm (bandpass) to maximise the amount of captured emission and to obtain reliable signal-to-noise ratios [253].

2.2.5 Chemical kinetics, curve fitting and model selection

Chemical kinetic reactions can be characterised by extracting the rate constant from the time dependence of reactant disappearance or product formation.

An example of first order kinetics data is a reaction where reactant A is converted to product B, and the rate of reaction (v) is proportional to the concentration of reactant A.

$$A \xrightarrow{k} B \tag{2.5}$$

$$v = \frac{\mathrm{d[B]}}{\mathrm{dt}} = -\frac{\mathrm{d[A]}}{\mathrm{dt}} = k[\mathrm{A}] \tag{2.6}$$

$$[\mathrm{A}] = [\mathrm{A}]_0 \mathrm{e}^{-kt} \tag{2.7}$$

The customary method for analysing first order kinetics data is to take the natural logarithms of both sides of Equation 2.7.

$$\ln[\mathrm{A}] = \ln[\mathrm{A}]_0 - kt \tag{2.8}$$

A plot of $\ln[A]$ vs. t is linear with a slope of $-k$ and intercept of $\ln[A]_0$. However, this analysis is problematic because the data are heteroskedastic, meaning that the error is not constant across the range. The error increases as $\ln[A]$ decreases, as shown in Figure 2.5 [266]. The consequence is that logarithms of values lower than the hypothetical ones deviate more than those higher than the hypothetical, leading to a steeper calculated slope, as can be seen from comparing the solid and dashed lines in Figure 2.5.

Rate constants can also be derived from measuring the formation of product B (Equation 2.9). For example, rate constants for unfolding reactions can be obtained by monitoring the formation of unfolded species.

$$[B] = [A]_0[1 - e^{-kt}] \tag{2.9}$$

The advantage of measuring product formation compared to measuring reactant disappearance is the higher signal-to-noise ratio when the reaction is reaching its equilibrium, thus enabling small changes in signal to be confidently detected. To obtain an equation which is linear with respect to t, Equation 2.9 can be solved for $[A]_0 - [B]$:

$$\ln([A]_0 - [B]) = \ln[A]_0 - kt \tag{2.10}$$

However, linear least squares fitting would be problematic, as $[A]_0$ must be known, and small errors in $[A]_0$ can cause large errors in k due to the increasing error as $\ln([A]_0 - [B])$ decreases [266].

Similar problems arise if reactant A does not react completely (Equations 2.11 & 2.12), for example when partially denaturing conditions are applied for studying protein unfolding and yields a mixture of folded species (the reactant) and unfolded species (the product). This leads to a non-zero amount of reactant at infinite time ($[A]_\infty$).

$$[A] - [A]_\infty = ([A]_0 - [A]_\infty)e^{-kt} \tag{2.11}$$

$$\ln([A] - [A]_\infty) = \ln([A]_0 - [A]_\infty) - kt \tag{2.12}$$

$[A]_0$ and $[A]_\infty$ must be known, and small errors in these values can cause large errors in k due to the increasing error as $\ln([A] - [A]_\infty)$ decreases [266].

One solution for dealing with heteroskedasticity is to use a weighted linear least-squares fit, where the values are weighted inversely to their variance. A problem with this method is that the weighting factor must be estimated from the data, which are unreliable [266].

The nonlinear least squares method is a more elegant way of fitting experimental data directly. Unlike the linear least squares method, there are no general analytical solutions available for non-linear least squares fitting. Instead, parameters are estimated through iterative calculations while minimising the sum of the squares of errors with minimisation algorithms such as the Levenberg-Marquardt algorithm. As illustrated in Figure 2.5, a comparison of linear least-squares fit against nonlinear least-squares fit of first order reaction kinetics data shows that better fitting of the simulated data and a more accurate rate constant are obtained with nonlinear least squares fitting [266].

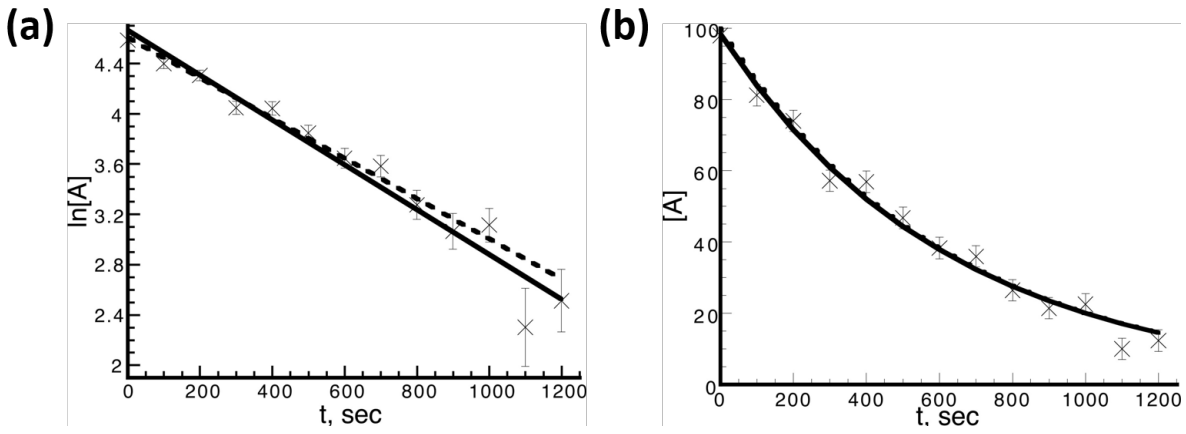


Figure 2.5: Comparison of linear vs. non-linear least squares analysis of kinetic data
(a) Plot of the best linear fit of Equation 2.8 (solid line) to simulated data (×), along with hypothetical values (dashed line). Error bars are shown to increase as [A] and ln[A] decrease. **(b)** Plot of the nonlinear least-squares fit of Equation 2.11 to simulated data (×), along with hypothetical values (dashed curve) and constant error bars representing the average error in [A]. Adapted from [266].

Two consecutive first-order reactions can be used for modelling more complex reaction schemes, for example the reaction of reactant X to form product Z via an intermediate Y, where each step in the reaction can be treated as a simple exponential. This scheme was employed in this chapter to describe the unfolding reaction of pSRII. The changes in [X], [Y] and [Z] over time are simulated in Figure 2.6a [267].

$$X \overset{k_1}{\rightarrow} Y \overset{k_2}{\rightarrow} Z \tag{2.13}$$

$$[X] = [X]_0 e^{-k_1 t} \tag{2.14}$$

$$[Y] = \frac{[X]_0 k_1}{(k_2 - k_1)}(e^{-k_1 t} - e^{-k_2 t}) \quad (2.15)$$

$$[Z] = [X]_0 \left\{1 + \frac{1}{(k_1 - k_2)}(k_2 e^{-k_1 t} - k_1 e^{-k_2 t})\right\} \quad (2.16)$$

where k_1 and k_2 are first-order rate constants.

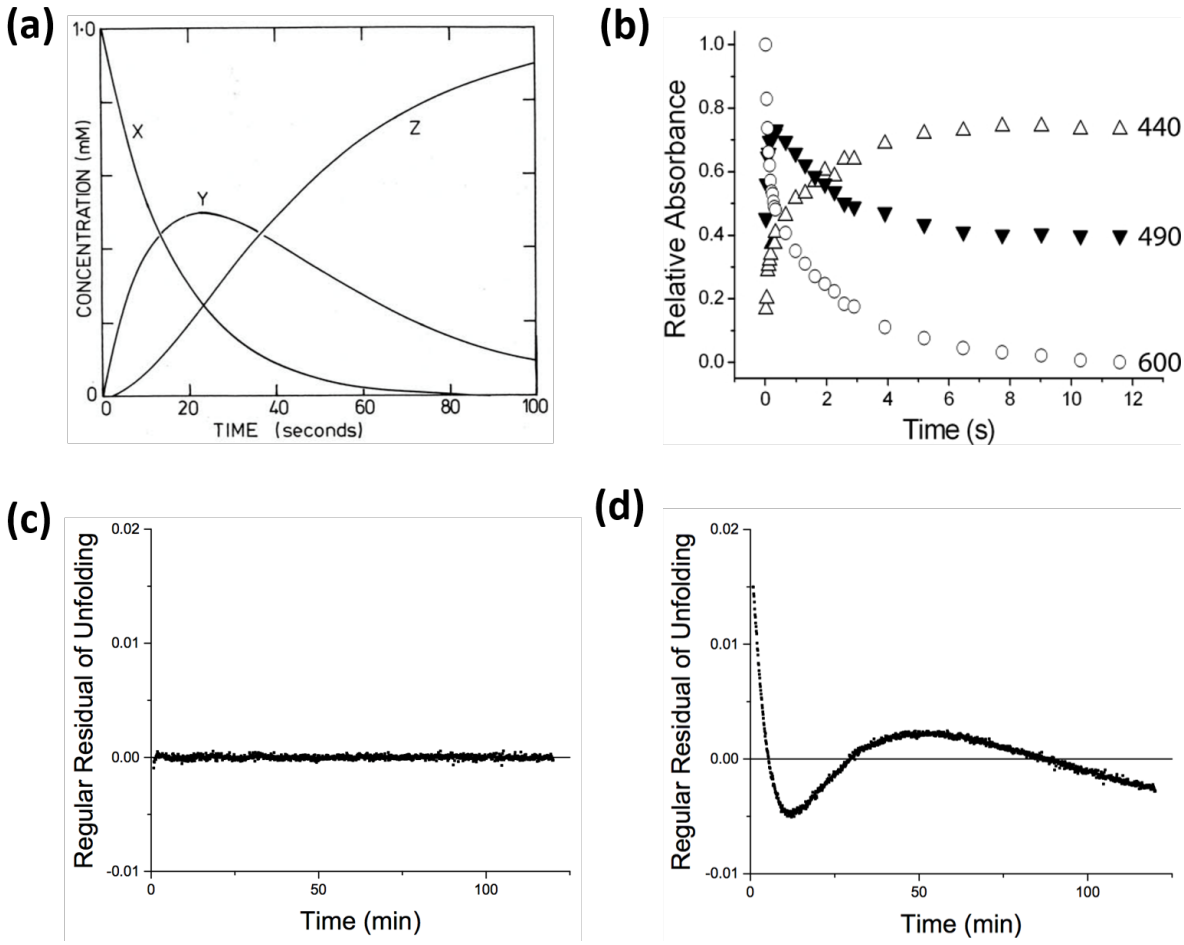


Figure 2.6: Estimating reaction rate constants from a two-step consecutive first-order reaction **(a)** Simulated changes in concentrations of X (reactant), Y (intermediate), and Z (product) following the scheme in Equation 2.13, with $k_1 = 6 \times 10^{-2}$ s^{-1}, $k_2 = 3 \times 10^{-2}$ s^{-1}, and $[X]_0 = 1$ mM. Retrieved from [267]. **(b)** Changes at different wavelengths over time in the absorption spectra of bacteriorhodopsin upon addition of 0.882 X_{SDS}: 600 nm (open circles), 490 nm (filled triangles), and 440 nm (open triangles). Retrieved from [116]. **(c–d)** Residuals from curve-fitting of absorbance at 600 nm with a double exponential equation **(c)** or with a single exponential equation **(d)** for bacteriorhodopsin unfolding in 0.73 X_{SDS}. Retrieved from [118].

The unfolding of bacteriorhodopsin has been modelled with the double first-order reaction scheme [116,118]. Upon mixing bacteriorhodopsin with denaturing concentrations of SDS, the native absorption band at 560 nm broadens and shifts to 600 nm (bR_{600}) within 16 ms.

The decay of bR_{600} coincides with the appearance of a band at 490 nm (bR_{490}), which is arises from retinal isomerisation to predominantly the 9-*cis* form. bR_{490} subsequently decays to a band at 440 nm (bO_{440}), an unfolded form of bacteriorhodopsin in which the retinal remains covalently-bound but most native protein-retinal interactions are lost. The retinal PSB is hydrolysed, and bO_{440} decays to bO_{390}. This scheme [116] is summarised as:

$$\mathrm{bR_{560}} \rightleftharpoons \mathrm{bR_{600}} \rightleftharpoons \mathrm{bR_{490}} \rightleftharpoons \mathrm{bO_{440}} \rightarrow \mathrm{bO_{390}} + \mathrm{retinal} \tag{2.17}$$

and can be further simplified as [118]:

$$\mathrm{bR_{600}} \underset{k_\mathrm{f}}{\overset{k_\mathrm{u}}{\rightleftarrows}} \mathrm{bO_{440}} \overset{k_\mathrm{h}}{\rightarrow} \mathrm{bO_{390}} + \mathrm{retinal} \tag{2.18}$$

Under unfolding conditions ($k_\mathrm{u} \gg k_\mathrm{f}$), the equilibrium between bR_{600} and bO_{440} can be simplified as an irreversible reaction to yield a double first-order reaction scheme. The unfolding of bacteriorhodopsin, monitored by the change in absorbance at 600 nm, consists of two kinetic phases (Figure 2.6c–d). Global analysis of the spectra performed using singular value decomposition showed that the data were best fitted to a sequential reaction model with three species corresponding to bR_{600}, bR_{490} and bO_{440}, respectively [116].

Choosing an appropriate fitting model is paramount to extracting accurate rate constants and making correct interpretations from kinetics data. A good model should balance the competing objectives of maximising both goodness-of-fit (conformation of fitted model to the data) and parsimony (model simplicity). Under-fitting excludes key variables or effects and induces bias. Over-fitting unnecessarily complicates the model by including extra variables.

The Akaike Information Criterion (AIC) [268] estimates the relative quality of different statistical models for a given set of data in order to provide a means for model selection. AIC is an implementation of Occam's razor, in which parsimony or simplicity is balanced against goodness-of-fit.

$$\mathrm{AIC} = 2\mathrm{q} - 2\log \mathrm{L} \tag{2.19}$$

where q is the number of model parameters and L is the likelihood (probability of the observed data coming from the estimated model). Given any two estimated models, the model with a smaller (or more negative) AIC value is more desirable.

One caveat with model comparison methods is that information about the quality of each model is not provided, meaning it is possible that none of the models fit the data adequately. Hence, model-based statistical analyses (i.e. testing a null hypothesis) are still required to assess the quality of a chosen model.

2.3 Results

2.3.1 Denaturant and additives screening identify SDS as the most suitable denaturant for unfolding pSRII

Characterisation of the denatured states and the unfolding and refolding pathways require maximal and reversible unfolding of the protein with minimal aggregation. To identify the most suitable denaturant for unfolding pSRII, the extent of protein unfolding was determined by monitoring changes in the UV/vis absorbance of the retinal chromophore. The results are summarised in Figure 2.7 and Table 2.1. The exquisite sensitivity of the retinal UV/vis absorbance fine structure enables facile monitoring of structural changes in the retinal binding pocket and detection of unfolding intermediates. In pSRII, all-*trans* retinal is covalently linked to Lys-205 via a PSB (Figure 2.7a) and gives rise to a main absorbance peak at 498 nm, with vibronic side bands at 460, 420 and 370 nm (Figure 2.7b) [224]. Retinal PSB devoid of native protein interactions absorbs at 440 nm (SR_{440}) [269]. Schiff base hydrolysis leads to the release of free all-*trans* retinal, which absorbs at 390 nm (SO_{390}) [240], from the sensory opsin apo-protein.

pSRII was subjected to denaturation for one day at room temperature in the presence of different denaturants (Figures 2.7c–e). These included chaotropic denaturants urea and guanidium hydrochloride (GdmCl), trifluoroacetic acid (TFA), 2,2,2-trifluoroethanol (TFE), anionic detergents N-lauroylsarcosine and sodium dodecyl sulphate (SDS), cationic detergent cetyltrimethylammonium bromide (CTAB), zwitterionic detergent lauryldimethylamine-N-oxide (LDAO), and non-ionic detergents polyoxyethylene (20) sorbitan monolaurate (Tween-20) and polyoxyethylene octyl phenyl ether (Triton X-100).

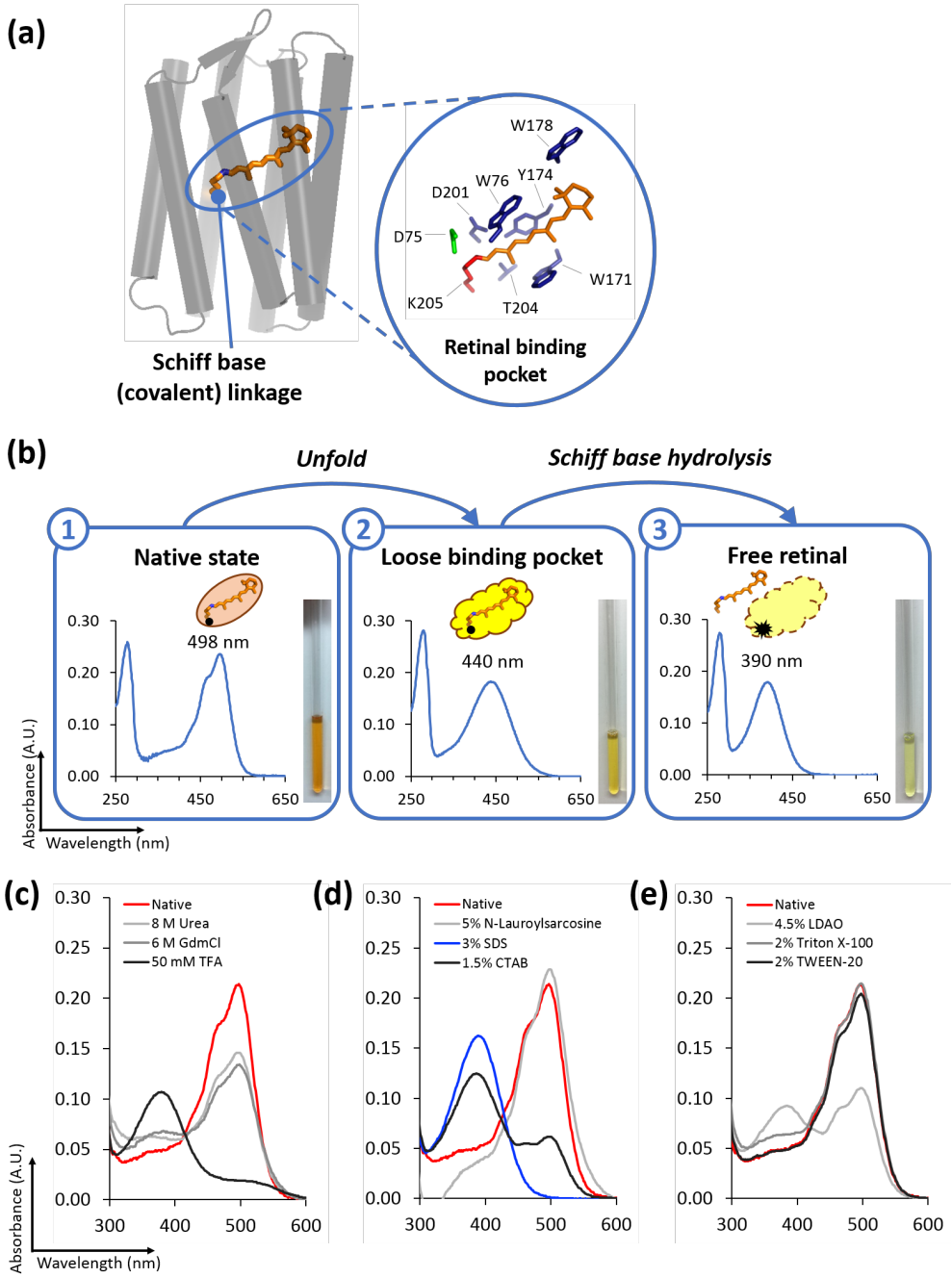


Figure 2.7: Denaturant screening for unfolding pSRII
(a) Retinal binding pocket of pSRII, where the all-*trans* retinal chromophore is covalently linked to Lys-205 via a PSB. **(b)** Changes in the UV/vis absorbance of the retinal chromophore during denaturation of pSRII. The insets show schematic representations of the binding pocket and the corresponding photographs of 90 µM pSRII in different spectral states. **(c–e)** Loss of retinal binding, determined by monitoring changes to the retinal absorbance peak by UV/vis spectroscopy, is taken as the hallmark of pSRII unfolding. All spectra are scaled to the 280-nm peak of the native spectrum (red). Only the 300- to 600-nm spectral range is displayed to illustrate changes in the retinal absorbance peak. pSRII was exposed to denaturation for one day in water-soluble denaturants **(c)**, anionic and cationic detergents **(d)**, and zwitterionic and non-ionic detergents **(e)**. Only SDS, an anionic detergent, led to complete loss of native retinal-protein contacts (λ_{max} = 390 nm).

	Denaturant	Example	Ref.	Absorbance	Visual Inspection	Conclusion
Water-soluble denaturants	8 M Urea	Bovine rhodopsin	[270]	No change (Native-like major peak at 498 nm with side bands at 460 nm, 420 nm and 370 nm)	Orange; clear	Does not unfold
	6 M GdmCl	Bovine rhodopsin	[270]	No change	Orange; clear	Does not unfold
	50% (v/v) TFE	KcsA potassium channel	[112,169]	N/A (aggregation observable by eye)	Pale yellow; cloudy	Unfolds but aggregates
	50 mM TFA	Bovine rhodopsin	[270]	Major peak at 390 nm; minor peak at 498 nm	Bright yellow; clear	Unfolds with slow kinetics
Anionic detergents	3% (w/v) SDS (0.980 X_{SDS})	Bovine rhodopsin; Bacteriorhodopsin	[107,270]	Single peak at 390 nm	Pale yellow; clear	Unfolds
	5% (w/v) N-lauroylsarcosine (0.988 $X_{\text{N-lauroylsarcosine}}$)	Olfactory receptor OR5	[271]	No change	Orange; clear	Does not unfold
Cationic detergent	1.5% (w/v) CTAB (0.952 X_{CTAB})	Cytochrome c oxidase	[272]	Major peak at 390 nm; minor peak at 498 nm	Pale yellow; clear	Unfolds with slow kinetics
Zwitterionic detergent	4.5% (w/v) LDAO (0.990 X_{LDAO})	Prokaryotic and eukaryotic transporters	[273]	Major peak at 498 nm; minor peak at 390 nm	Yellow; clear	Unfolds with very slow kinetics
Non-ionic detergents	2% (v/v) Triton X-100 (0.941 $X_{\text{Triton X-100}}$)	5-HT_{1A} receptor	[274]	No change	Orange; clear	Does not unfold
	2% (v/v) Tween-20 (0.896 $X_{\text{Tween-20}}$)	Membrane protein extraction	[275]	No change	Orange; clear	Does not unfold

Table 2.1: Effects of different denaturants on 7 µM pSRII after a one-day incubation at room temperature
TFE, 2,2,2-trifluoroethanol; TFA, trifluoroacetic acid; SDS, sodium dodecyl sulphate; CTAB, cetyltrimethylammonium bromide; LDAO, lauryldimethylamine-N-oxide; Triton X-100, polyoxyethylene octyl phenyl ether; Tween-20, polyoxyethylene (20) sorbitan monolaurate.

Amongst the denaturants screened, only SDS yielded complete unfolding and Schiff base hydrolysis, with only SO_{390} being observed (Figure 2.7d). Like bacteriorhodopsin [276], pSRII cannot be denatured by urea and GdmCl, whereas bovine rhodopsin is susceptible to both denaturants [270]. TFA led to the formation of SR_{440} which slowly decayed to form SO_{390} over the course of two weeks, indicating that the rate of Schiff base hydrolysis is reduced considerably in an acidic environment. TFE led to aggregation which was observable by naked eye, and the sample was therefore not analysed by UV/vis spectroscopy. pSRII cannot be denatured by LDAO, even though it is a relatively harsh detergent, and it has been proposed that only ~20% of membrane proteins are resistant to its denaturing effects [277].

Aggregation was assessed by SDS-PAGE without boiling of the samples. Native loading dye (see Materials and Methods) was used so that no additional SDS was introduced from the sample buffer. Figure 2.8a shows a silver-stained SDS-PAGE gel for pSRII in 0%, 0.5% and 3% SDS in the presence of 0.1% c7-DHPC for 0–72 h. Native pSRII in 0.1% c7-DHPC ran as a major monomeric band and a minor dimeric band. Sedimentation velocity analytical ultracentrifugation of 4.9 μM pSRII in 0.1% c7-DHPC confirmed that the monomeric species accounts for more than 90% of the total loading sample, while the dimeric species accounts for about 4% [278], consistent with the relative intensities of the two bands on SDS-PAGE gel. The amount of oligomerisation increased over prolonged unfolding time (Figure 2.8a) and correlated with the loss of α-helical structure and increase in β-sheet structure over time (Figure 2.8c). While dimers and higher-order oligomeric species were observed in both 0.5% and 3% SDS, the amount of higher oligomeric states was lower in 3% SDS (Figures 2.8a–b). This suggests that 0.5% SDS induced slow or partial unfolding, and the exposed protein surfaces led to preferential formation of higher oligomers. pSRII remained predominantly monomeric, and the amount of aggregation did not show noticeable increase after 2 h in SDS, which is the typical timeframe for the biophysics and kinetics studies in this chapter. The amounts of dimers and higher-order oligomers steadily increased from 24 h onward in both 0.5% and 3% SDS (Figure 2.8a).

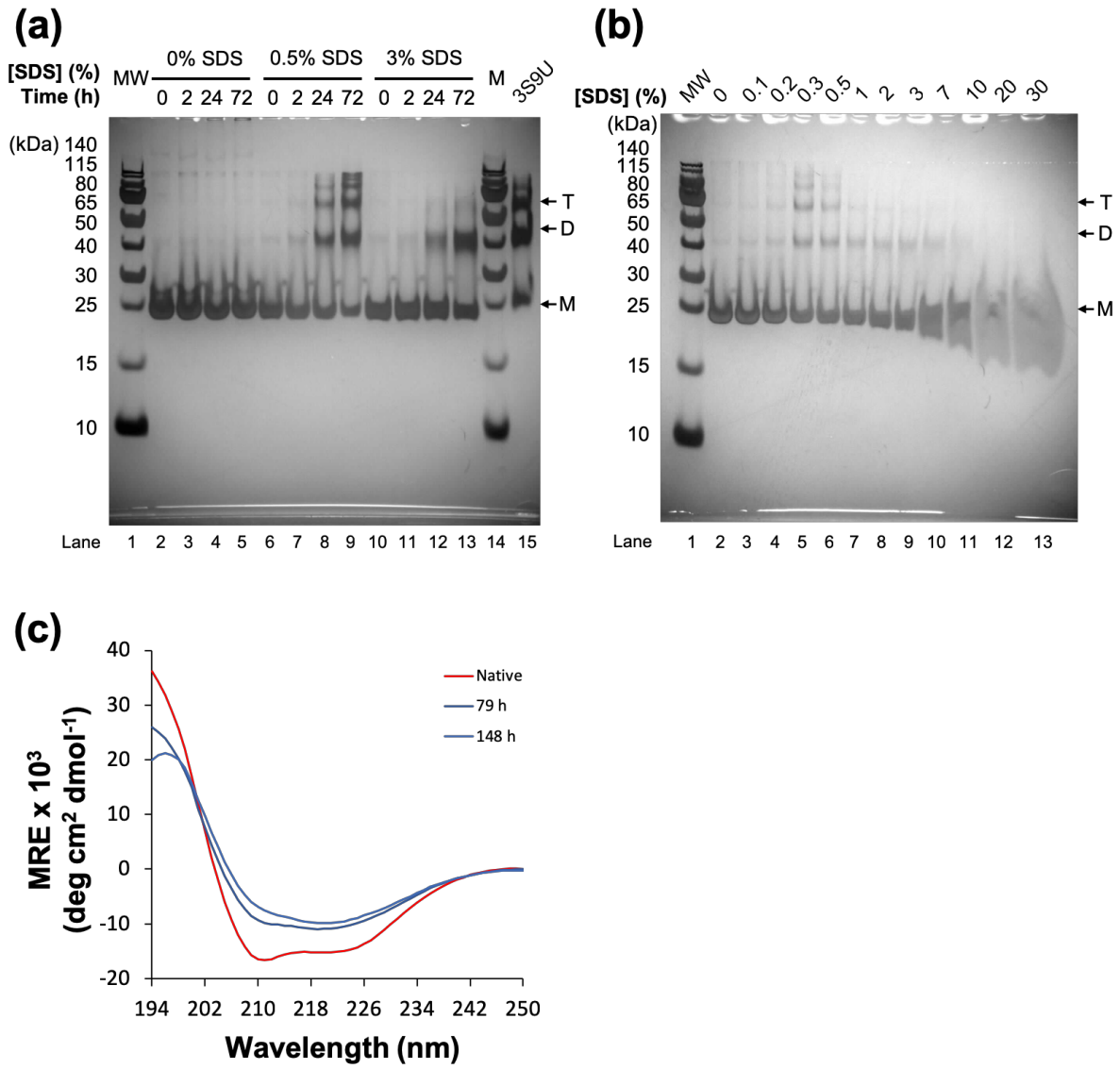


Figure 2.8: Detection of aggregates for pSRII unfolded in SDS + 0.1% c7-DHPC at pH 6.0
(a) Silver-stained 12% SDS-PAGE gel showing increased aggregation over prolonged unfolding time of pSRII in 0% (lanes 2–5), 0.5% (lanes 6–9) and 3% SDS (lanes 10–13) at pH 6.0 for 0 min, 2 h, 24 h and 72 h at 25 °C. Lanes 1 and 14, molecular weight marker; lane 15, pSRII boiled in 3% SDS + 9 M urea (3S9U) for 5 min at 95 °C. **(b)** Silver-stained 12% SDS-PAGE gel of pSRII unfolded in different concentrations of SDS + 0.1% c7-DHPC at pH 6.0 for 14 h at 25 °C. Lane 1, molecular weight marker; lane 2, pSRII; lanes 3–13, pSRII in 0.1–30% SDS. In panels a and b, the arrows indicate the expected migration of monomeric (M), dimeric (D) and trimeric (T) pSRII. **(c)** CD spectra of 7 µM (0.2 mg/ml) native pSRII and pSRII unfolded in 0.998 X_{SDS} (30% SDS in 0.1% c7-DHPC) at pH 6.0 for up to 148 h, showing loss of α-helical structure and gain of β-sheet structure over time.

The oligomerised species were hypothesised to represent off-pathway species that cannot be refolded to restore the native chromophore. To optimise the unfolding condition, different

additives were explored with the goal of eliminating or reducing aggregation, and the results are summarised in Table 2.2. Protein aggregation was monitored by SDS-PAGE and the extent of protein unfolding was determined by UV/vis spectroscopy (Figure 2.9). Additives commonly used for reducing protein aggregation [279] were added to unfolding buffers containing 3% SDS, including non-detergent sulfobetaines NDSB-201 (3-(1-Pyridinio)-1-propanesulfonate) and NDSB-256 (3-(Benzyldimethylammonio)propanesulfonate), L-arginine, polyethylene glycol PEG-8000, LDAO (Lauryldimethylamine-N-oxide) and urea. Different solvent conditions, including salt concentration (0–500 mM NaCl) and pH (pH 1.0–6.0, adjusted using TFA or phosphoric acid), were also explored.

Additive/Buffer	Final Conc.	Ref.	Aggregation after 1 day	Conclusion
NaCl	0–500 mM	[279]	Yes	More oligomeric species than in the absence of additive
PEG-8000	0–30 mg/ml	[279]	Yes	More oligomeric species than in the absence of additive
LDAO	0.18%	[280]	Yes	Similar amount of oligomeric species as in the absence of additive
NDSB-201	1 M	[279]	Yes	More oligomeric species than in the absence of additive
NDSB-256	1 M	[279]	Yes	More oligomeric species than in the absence of additive
L-arginine	0.5 M	[279]	No	pSRII did not unfold (see Figure 2.9d)
Urea	0–8 M	[270]	Yes	More oligomeric species than in the absence of additive
TFA	0–200 mM	[87,281,282]	Yes	Amount of oligomerisation depends on pH or [TFA]
Sodium phosphate, pH 2.0	50 mM	[283]	Yes	Less oligomeric species than at pH 6.0

Table 2.2: Aggregation of pSRII unfolded for one day at room temperature in 3% SDS + different additives or buffer conditions.

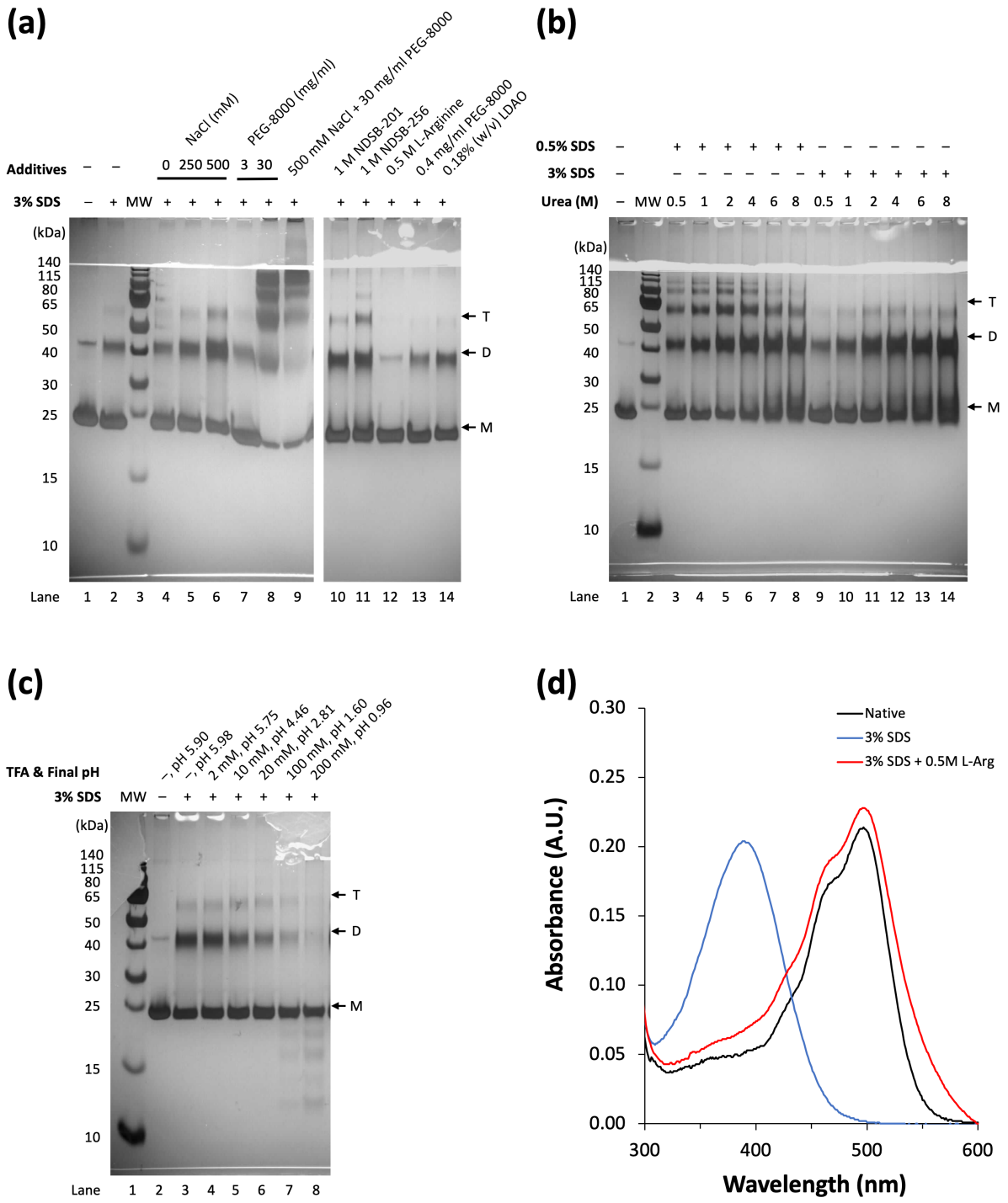


Figure 2.9: Denaturation of pSRII in SDS leads to some extent of aggregation over time
(a) Silver-stained 12% SDS-PAGE gel of pSRII unfolded for 16 h in SDS in the presence of different additives, in attempt to decrease oligomerisation. **(b)** SDS-PAGE of pSRII unfolded for 14.5 h in SDS + urea. **(c)** SDS-PAGE of pSRII unfolded in 3% SDS + varied concentrations of TFA for 62.5 h shows pH-dependent unfolding, aggregation and degradation. In panels a–c, the arrows indicate the expected migration of monomeric (M), dimeric (D) and trimeric (T) pSRII. **(d)** UV/vis spectra of the retinal chromophore of pSRII in the native condition (black trace), 3% SDS (blue trace) and 3% SDS + 0.5 M L-arginine (red trace) show that L-arginine has prevented the protein from unfolding.

Non-detergent sulfobetaines (NDSBs) are amphiphilic molecules with a short hydrophobic alkyl chain and a zwitterionic sulfobetaine group. NDSBs are not detergents since their short hydrophobic tails preclude micelle formation. Concentrations of NDSBs in the 1–2 M range have been used for refolding chemically- and thermally-denatured proteins. Renaturation studies on hen egg white lysozyme and bacterial β-D-galactosidase suggest that NDSBs interact with early folding intermediates and weaken abortive hydrophobic interactions which would have promoted misfolding and aggregation. The mechanism is proposed to be aromatic stacking interactions of the cyclic moiety of NDSBs with exposed aromatic groups of denatured proteins [284]. NDSBs are attractive addititves for SDS-mediated unfolding reactions, since they mediate a balance between protein folding and aggregation whilst preventing non-specific interactions. pSRII denatured with 3% SDS in the presence of 1 M NDSB-201 or 1 M NDSB-256 (Figure 2.9a, lanes 10–11) both led to the formation of more dimer and higher oligomers, with more pronounced aggregation observed in the presence of NDSB-256. This indicates that neither NDSBs were effective in reducing the aggregation of unfolded pSRII, causing worse shielding of exposed aggregation-prone surfaces.

Arginine hydrochloride (L-Arg) is commonly used as an aggregation suppressor in protein refolding [279,285]. L-Arg increases the solubility of aggregation-prone unfolded species and intermediates of hen egg white lysozyme [286]. Computational studies show that L-Arg stabilises partially unfolded intermediates by interacting with aromatic residues through cation−π interaction and also with charged residues through salt bridges. The self-interaction of L-Arg leads to cluster formation, crowding out protein-protein interactions to eliminate aggregation [287]. SDS-PAGE of pSRII denatured in 3% SDS + 0.5 M L-Arg shows minimal amount of aggregation, but the protein has not unfolded (λ_{max} = 498 nm; Figure 2.9d). L-Arg's mechanism of action is thought to be complex, as its effects are concentration- and protein-dependent. It is speculated that L-Arg simultaneously affects the stability of the native state and the unfolded states, as demonstrated with the protein KMP-11 [288]. For the case of pSRII, L-Arg might have led to an overall greater destabilisation of the unfolded states.

Polyethylene glycols (PEGs) are polymers with a variety of polymer chain lengths and molecular weights. Such water-soluble biopolymers could lead to either accelerated or

reduced protein aggregation through a complex interplay of macromolecular crowding (decreasing the effective volume of a protein and stabilising more compact conformations) and non-specific protein-polymer interactions (electrostatic, hydrophobic and/or van der Waals) [289]. 1D ^{1}H NMR and WATERGATE-TOCSY NMR data indicate that PEG-4000 binds in a hydrophobic cleft on the surface of hen egg-white lysozyme formed by Arg-61, Trp-62, Trp-63, Arg-73, Lys-96 and Asp-101 [290]. pSRII showed more pronounced aggregation in the presence of increasing concentrations of PEG-8000 (0.4, 3 and 30 mg/ml) with 3% SDS (Figure 2.9a lanes 7, 8 & 13). Since the properties and effects of PEGs on aggregation and protein structure are highly dependent on their molecular weights [291], it is unknown whether using PEGs of other molecular weights could reduce the aggregation of pSRII more effectively than PEG-8000.

As SDS is an anionic detergent, SDS concentration-dependent oligomerisation during pSRII denaturation might be mediated by interactions between detergent micelle surfaces [292]. Slight adjustments of surface charges can be achieved by using mixed micelles. Zwitterionic surfactants can have strong interactions with anionic surfactants in solution [293]. The zwitterionic detergent LDAO was chosen, given its similar chain length as SDS. Size exclusion chromatography of pSRII shows a double peak corresponding to monomer and dimer in DDM, whereas only monomer peak is observed in 0.1% LDAO [280], suggesting possible compatibility of LDAO with reducing pSRII oligomerisation. Moreover, LDAO has been shown to prevent aggregation and allow functional folding of unfolded MamC, a magnetosome-associated membrane protein from the magnetotactic bacterium *Magnetococcus marinus* MC-1 [294]. pSRII denatured in 3% SDS + 0.18% LDAO showed similar oligomerisation patterns on the SDS-PAGE gel (Figure 2.9a, lane 14), suggesting that LDAO was insufficient to perturb surface charges at the concentration tested. Further tests with higher LDAO concentrations were not attempted given potential difficulties in reliably replicating the properties of mixed micelles containing three detergents (c7-DHPC, SDS and LDAO) in further biophysical studies.

A combination of SDS and urea (3% SDS + 8 M urea) is shown to be the most effective condition for denaturing bovine rhodopsin [270], where the amphiphilic SDS shields aggregation-prone surfaces which are exposed by urea. Fluorescence measurements using the

probes pyrene-3-carboxaldehyde (PCA) localised to the surface of SDS micelles and diphenylbutadiene (DPB) solubilised in SDS micellar phase show that urea interacts with the polar headgroups and hydrocarbon chains of SDS micelles, leading to a more rigid microenvironment and displacement of some water molecules from the solvation layer of the surfactant hydrophobic chains in the micelle [295]. pSRII denatured in different concentrations of urea (0.5–8 M) in the presence of 0.5% or 3% SDS (Figure 2.9b) showed more pronounced oligomerisation with increasing urea concentration, with more oligomerisation being observed with 0.5% SDS than with 3% SDS. These results suggest that structural perturbation to SDS micelles by urea is inversely dependent on SDS concentration and leads to more unfavourable exposure of aggregation-inducing surfaces on pSRII.

Changing the salt concentration perturbs surface charges and the strength of electrostatic interactions. 50 mM NaCl is already present in the unfolding buffer without additives. Denatured pSRII aggregated faster in unfolding buffers with higher salt concentrations (Figure 2.9a lanes 2 & 4–6). Hence, increasing the ionic strength was unfavourable for maintaining the monomeric state of SDS-denatured pSRII.

TFA is a harsh organic acid which is routinely used for solubilising peptides [296]. SDS in the presence of different concentrations of TFA at different pHs were also explored (Figure 2.9c). Changing the pH affects both the surface charge distribution and the net charge of the protein, hence modulating electrostatic interactions which could have contributed towards aggregation. As pH decreases with increasing TFA concentrations, oligomeric bands became fainter, indicating that lower pH slows down aggregation. However, protein degradation was observed in SDS at pH 1.60 and pH 0.96, indicating that further optimisation is required to balance oligomerisation and acid-mediated proteolysis.

Controls against possible protein modifications by TFA, such as trifluoroacetylation [296], were accounted for by unfolding pSRII at pH 2.0, adjusted using phosphoric acid, both in the absence and presence of SDS (Figure 2.10). The formation of oligomers in SDS at pH 2.0 showed similar dependence on time and SDS concentration as in SDS at pH 6.0.

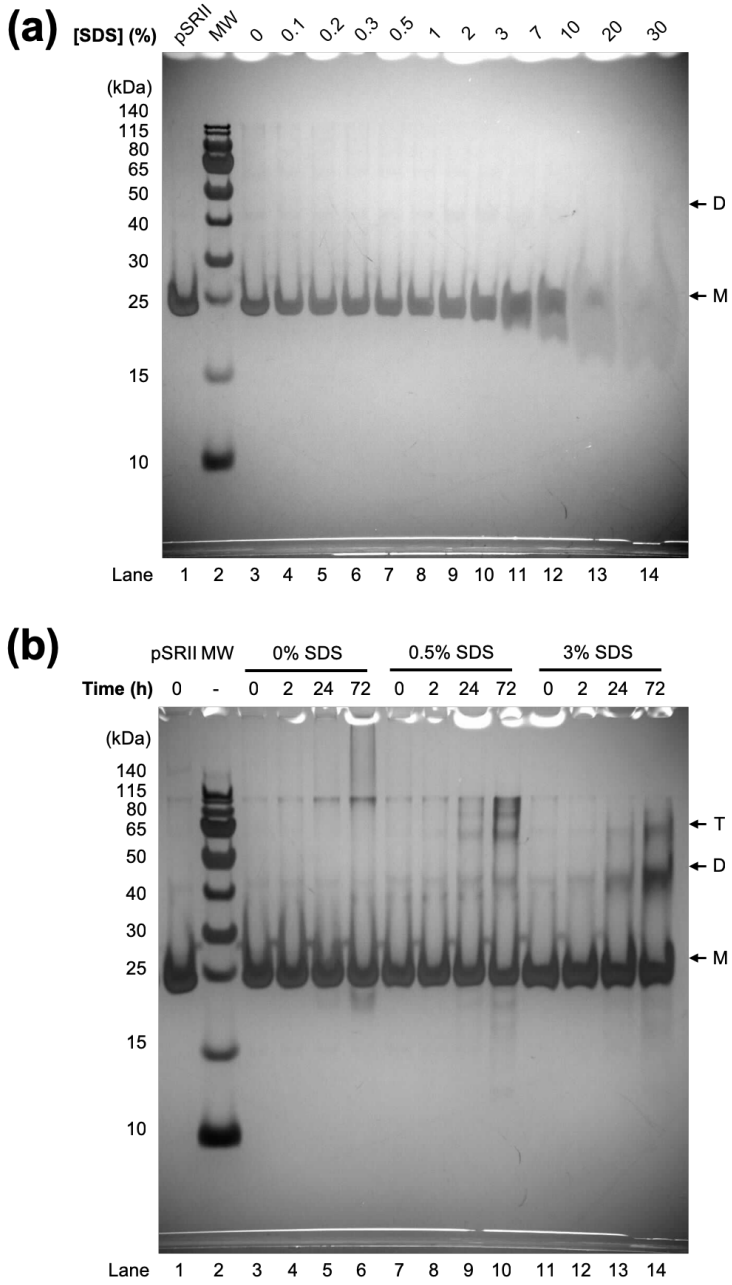


Figure 2.10: Detection of aggregates for pSRII unfolded in SDS at pH 2.0
(a) SDS-PAGE of pSRII unfolded in different concentrations of SDS + 0.1% c7-DHPC at pH 2.0 for 14 h at 25 °C. Lane 1, pSRII at pH 6.0; lane 2, molecular weight marker; lanes 3–14, pSRII in 0–30% SDS at pH 2.0. **(b)** SDS-PAGE showing time-dependent aggregation of pSRII unfolded in SDS at pH 2.0. Lane 1, pSRII at pH 6.0; lane 2, molecular weight marker; lanes 3–14, pSRII unfolded for 0 min, 2 h, 24 h and 72 h in 0% (lanes 2–5), 0.5% (lanes 6–9) and 3% SDS (lanes 10–13), pH 2.0. In both panels, the arrows indicate the expected migration of monomeric (M), dimeric (D) and trimeric (T) pSRII.

In summary, despite extensive efforts, none of the tested additives and buffer conditions offered decreased oligomerisation without compromising the unfolding rate (see Figure 2.9d).

Results interpretation was complicated by unknown effects of different additives on pSRII and SDS/c7-DHPC mixed micelles. SDS-denaturation of pSRII at pH 6.0 and at pH 2.0 were identified as conditions which yield maximal unfolding and the least amount of aggregation. Further biophysical studies were therefore pursued in different SDS concentrations at pH 6.0 and pH 2.0 to study structural changes and kinetics of pSRII unfolding.

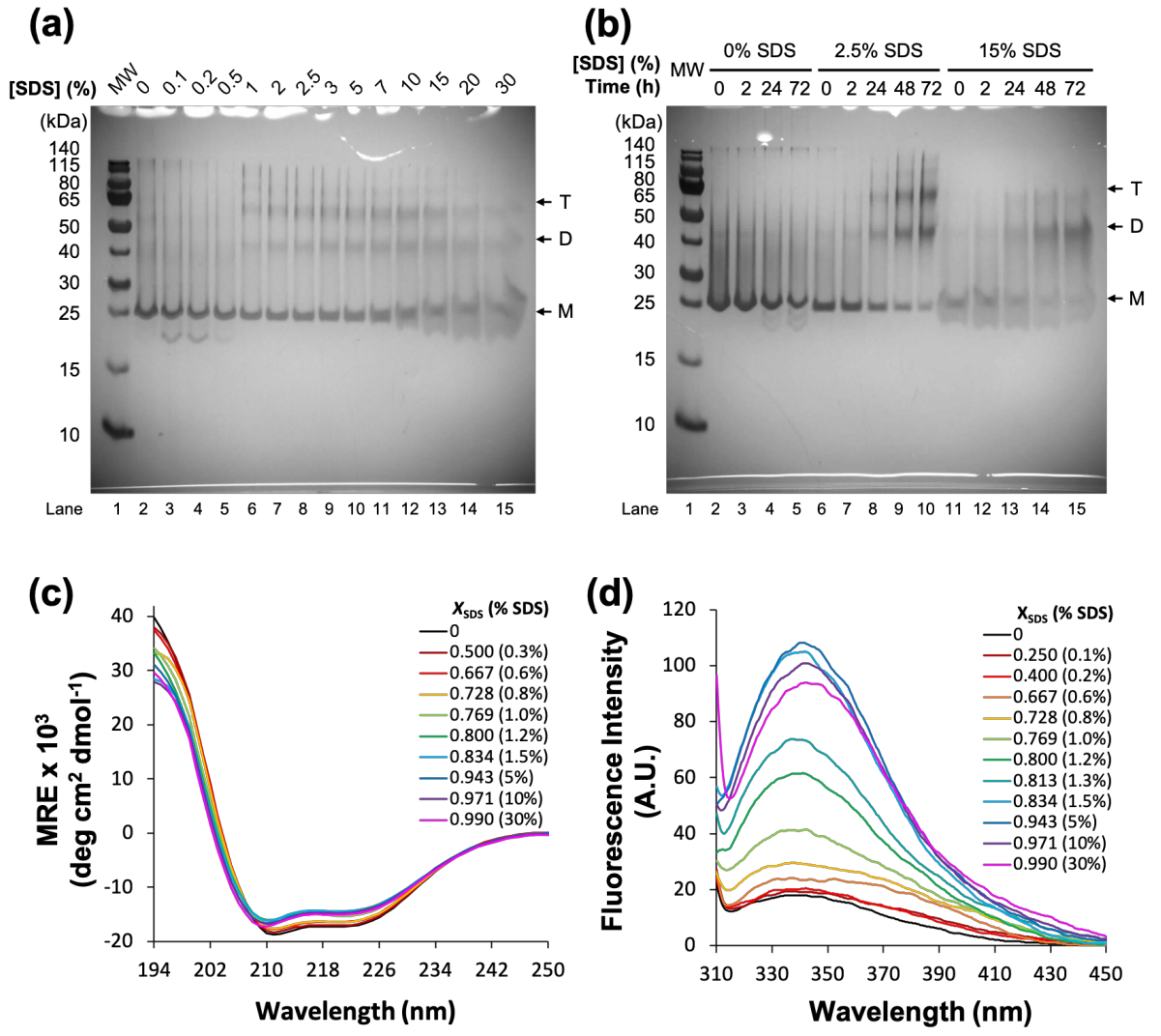


Figure 2.11: pSRII unfolding in SDS + 0.5% c7-DHPC at pH 6.0
(a) Silver-stained 12% SDS-PAGE gel of pSRII unfolded in different concentrations of SDS + 0.5% c7-DHPC at pH 6.0 for 14 h at 25 °C. Lane 1, molecular weight marker; lane 2, pSRII; lanes 3–13, pSRII in 0.1–30% SDS. **(b)** Silver-stained 12% SDS-PAGE gel showing increased aggregation over prolonged unfolding times of pSRII in 0% (lanes 2–5), 2.5% (lanes 6–10) and 15% SDS (lanes 11–15) + 0.5% c7-DHPC at pH 6.0 for 0–72 h at 25 °C. In panels a–b, the arrows indicate the expected migration of monomeric (M), dimeric (D) and trimeric (T) pSRII. **(c–d)** CD spectra **(c)** and tryptophan fluorescence emission spectra **(d)** for 7 µM (0.2 mg/ml) pSRII in 0–0.990 X_{SDS} (0–30% SDS in 0.5% c7-DHPC).

2.3.2 Characterisation of denatured states: SDS at pH 6.0

Based on the combined criteria of the largest degree of denaturation and the least amount of aggregation, in the subsequent sections, structural properties of SDS-denatured pSRII were studied at a maximum of 2 h after initial exposure to SDS, thus ensuring complete unfolding and negligible amounts of aggregates accumulated across all conditions. All biophysics data in this chapter were correlated with the bulk mole fraction of SDS (X_{SDS}), as this parameter describes the local environment to which the protein is being exposed in terms of the amount of SDS available in solution relative to the total amount of detergent (SDS + c7-DHPC). Since the extent of membrane protein unfolding correlates with X_{SDS} [116], high X_{SDS} in low c7-DHPC concentration (0.1% c7-DHPC) was used in order to obtain maximally unfolded pSRII for biophysical characterisation of SDS-denatured states. The kinetics of unfolding and refolding reactions were studied in the presence of excess concentrations of SDS/c7-DHPC mixed micelles (SDS + 0.5% c7-DHPC) to obtain accurate rate constants while avoiding multiple micelle occupancy. Control experiments show that similar extents of aggregation, and similar secondary and tertiary structure changes were yielded for SDS-denatured pSRII in the presence of 0.1% c7-DHPC vs. 0.5% c7-DHPC (Figure 2.11).

2.3.2.1 Secondary structure changes

The loss of secondary structure in pSRII, unfolded for 2 h at pH 6.0 with 0–0.998 X_{SDS} (0–30% SDS in 0.1% c7-DHPC), was monitored by CD spectroscopy. Little change was observed, suggesting that α-helical content was largely preserved in SDS-denatured pSRII at pH 6.0 (Figure 2.12a). Similar changes in α-helical content were observed for pSRII unfolded in SDS + 0.5% c7-DHPC (Figure 2.11c). Changes in mean residue ellipticity (MRE) at 222 nm from folded pSRII can be divided into two distinct stages: X_{SDS} up to 0.980 (3% SDS in 0.1% c7-DHPC) led to a decrease in MRE magnitude by up to 12.6 ± 1.0%; further increase in X_{SDS} reversed this trend, with only 9.3 ± 1.6% decrease in MRE magnitude observed at 0.998 X_{SDS} (30% SDS in 0.1% c7-DHPC) (Figure 2.12b). This slight increase in helicity at high X_{SDS} could be attributed to SDS favouring the formation of non-native helical structures [297].

Changes in different secondary structure elements calculated using CDPro software [298] are shown in Table 2.3 and Figure 2.13. The fraction of regular helix (H(r)) showed significant

decrease in > 0.834 X_{SDS}. The fraction of distorted helix (H(d)) increased in > 0.893 X_{SDS}, while the content of other secondary structure elements (regular sheet, distorted sheet, turn and unordered) only showed slight increases in > 0.834 X_{SDS}.

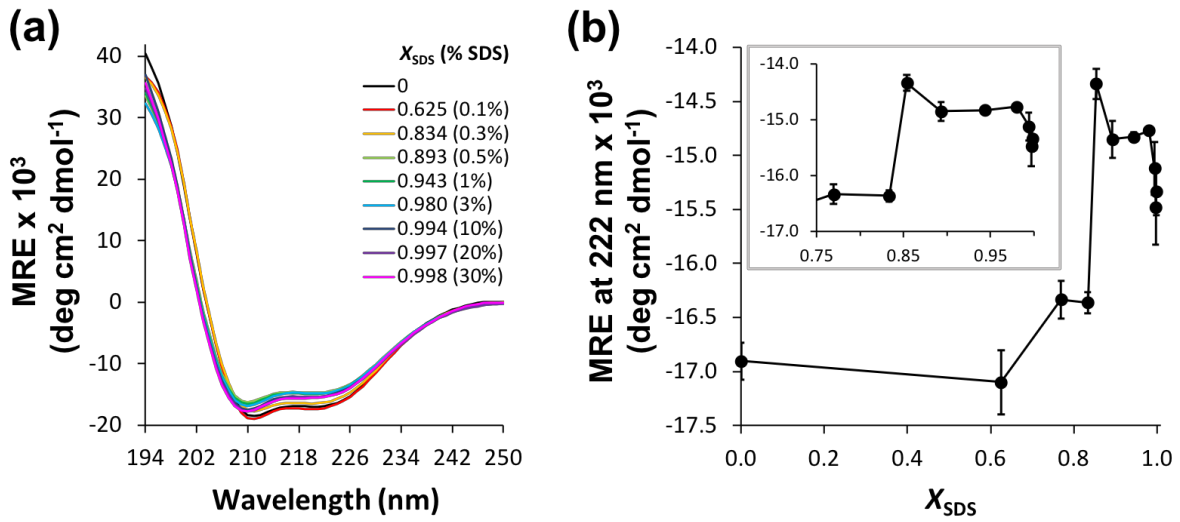


Figure 2.12: Secondary structure changes in SDS-denatured pSRII at pH 6.0
(a) CD spectra and **(b)** MRE at 222 nm for 7 µM (0.2 mg/ml) pSRII in 0–0.998 X_{SDS} (0–30% SDS in 0.1% c7-DHPC). Inset in panel b shows expanded view for the 0.75–1.00 X_{SDS} region. Error bars represent the standard deviation of three independent measurements.

X_{SDS}	H(r)	H(d)	S(r)	S(d)	T	U
0	47.3 ± 1.3%	17.0 ± 0.7%	0.6 ± 0.1%	3.1 ± 0.4%	10.4 ± 0.6%	21.6 ± 0.3%
0.625	47.5 ± 1.2%	17.4 ± 0.3%	0.4 ± 0.2%	2.7 ± 0.2%	10.6 ± 0.5%	21.4 ± 0.4%
0.834	45.6 ± 0.7%	17.4 ± 0.4%	0.8 ± 0.3%	3.3 ± 0.2%	11.6 ± 0.3%	21.3 ± 0.2%
0.893	38.1 ± 0.4%	18.6 ± 0.2%	2.2 ± 0.2%	3.9 ± 0.1%	13.9 ± 0.3%	23.2 ± 0.2%
0.943	38.7 ± 0.3%	19.0 ± 0.1%	1.7 ± 0.2%	3.9 ± 0.1%	13.6 ± 0.2%	23.0 ± 0.1%
0.980	38.9 ± 1.8%	19.1 ± 0.8%	1.4 ± 0.4%	4.0 ± 0.2%	13.6 ± 0.5%	23.0 ± 0.6%
0.994	40.7 ± 2.6%	20.3 ± 2.1%	2.4 ± 1.9%	3.0 ± 1.0%	8.9 ± 7.8%	24.5 ± 3.3%
0.997	41.1 ± 0.5%	20.0 ± 0.7%	1.5 ± 1.0%	3.6 ± 0.3%	12.1 ± 1.7%	21.7 ± 2.8%
0.998	40.3 ± 1.4%	19.6 ± 0.4%	0.8 ± 0.7%	3.6 ± 0.3%	13.1 ± 0.4%	22.5 ± 0.6%

Table 2.3: Deconvolution of CD spectra for pSRII unfolded for 2 h in SDS at pH 6.0
Deconvolution was performed using CDPro with CONTINLL algorithm and basis set no. 10 (contains soluble and membrane proteins). H(r), regular helix; H(d), distorted helix; S(r), regular strand; S(d), distorted strand; T, turns; U, unordered. Errors are reported as standard deviations of three independent measurements.

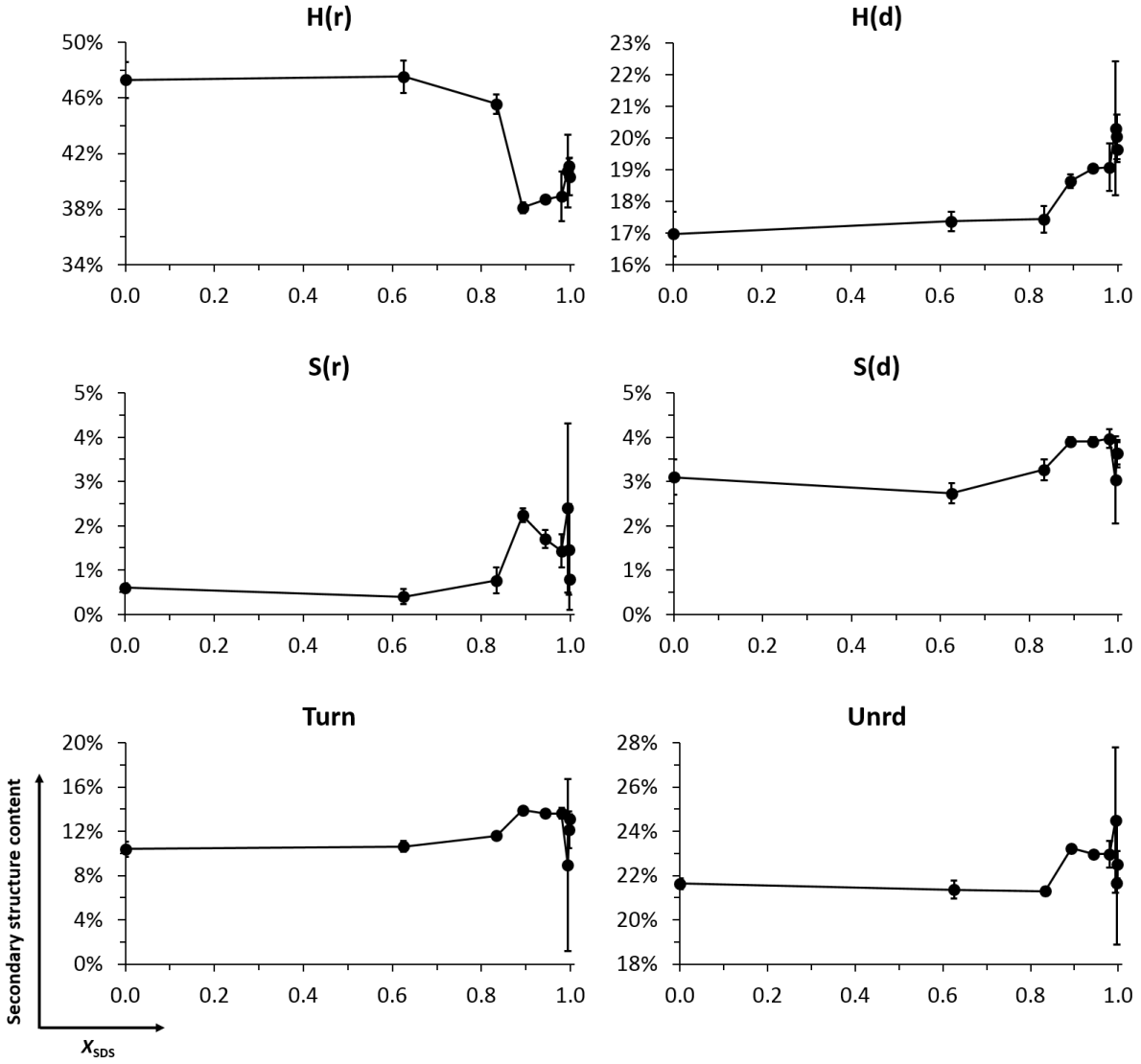


Figure 2.13: Changes in secondary structure composition of pSRII in SDS at pH 6.0
CD spectra were deconvoluted using CDPro with CONTINLL algorithm and basis set no. 10 (contains soluble and membrane proteins). H(r), regular helix; H(d), distorted helix; S(r), regular strand; S(d), distorted strand; T, turns; U, unordered. Errors are reported as standard deviations of three independent measurements.

2.3.2.2 Changes in the retinal binding pocket

UV/vis spectra of the retinal chromophore were recorded to monitor structural changes at the retinal binding pocket. In low X_{SDS} (≤ 0.893 X_{SDS}; 0.5% SDS in 0.1% c7-DHPC), an isosbestic point at ~420 nm was observed. A spectral intermediate absorbing at 440 nm (SR_{440}), characteristic of a PSB devoid of native retinal-protein contacts, appeared in ≥ 0.943 X_{SDS} (1% SDS in 0.1% c7-DHPC). Schiff base hydrolysis occurs over the course of 2 h, leading to the release of free all-*trans* retinal, which absorbs at 390 nm (SO_{390}) [240], yielding the sensory opsin apo-protein (Figure 2.14a).

To accurately determine the unfolding transition, equilibrium unfolding measurements were performed by unfolding pSRII in SDS in the presence of 0.5% c7-DHPC and monitoring the loss of absorbance at 498 nm after 10 min (Figure 2.14b). 0.5% c7-DHPC was used to avoid multiple micelle occupancy by ensuring an excess molar ratio of detergent micelles to protein across all X_{SDS}. The transition midpoint was around 0.81 X_{SDS}, higher than those reported for bacteriorhodopsin [116,299], indicating that pSRII in c7-DHPC is more stable than bacteriorhodopsin.

Each UV/vis spectrum was deconvoluted by nonlinear least squares fitting of the chromophore absorbance peaks to a model of Gaussian peaks (Figure 2.15). Gaussian-fitted λ_{max} of dead-time spectra in different X_{SDS} (dead-time = 25 s) revealed X_{SDS}-dependent red shift of the chromophore absorbance peak which closely matches the increase in unfolded protein fraction obtained from equilibrium unfolding measurements (Figure 2.14b). Since the λ_{max} values of protonated retinylidene model Schiff bases are sensitive to the proximity of the counteranion [300], spectral red shift of the retinal chromophore could suggest an increase in the separation of the Schiff base counterion, Asp-75, from the PSB upon denaturation of pSRII in SDS. The fitted λ_{max} remained constant beyond 0.943 X_{SDS} (5% SDS in 0.5% c7-DHPC), suggesting that further increases in X_{SDS} did not yield further separation of Asp-75 from the PSB, and the binding pocket is already fully opened.

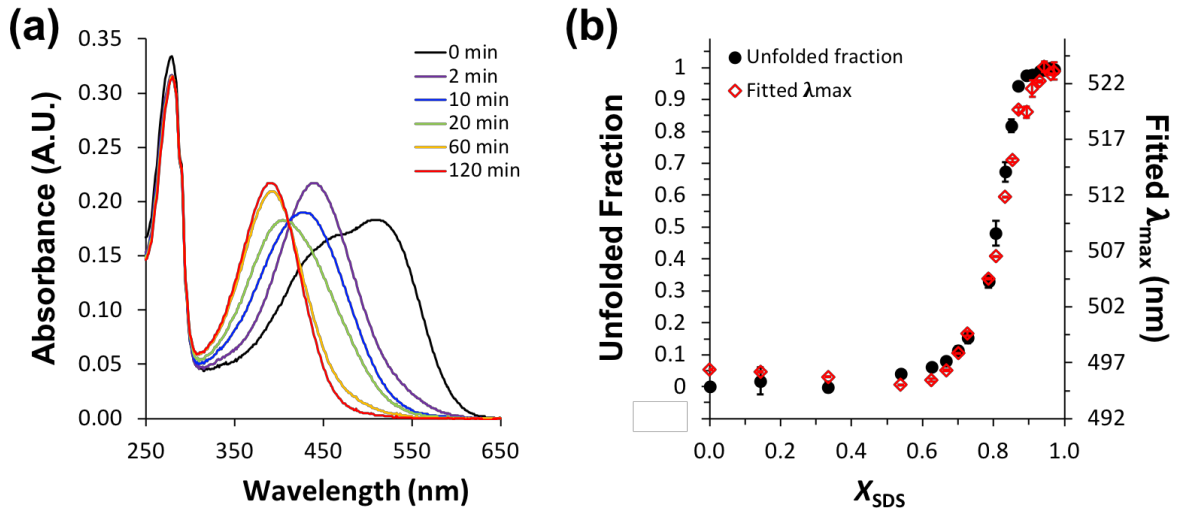


Figure 2.14: Loosening of the retinal binding pocket in SDS-denatured pSRII at pH 6.0
(a) Time-resolved UV/vis spectra of pSRII unfolded in 0.980 X_{SDS} (3% SDS in 0.1% c7-DHPC). **(b)** SDS-induced equilibrium unfolding of pSRII in the presence of 0.5% c7-DHPC monitored by UV/vis spectroscopy. λ_{max} of the retinal chromophore at t = 0 min, obtained by nonlinear least squares fitting, is also plotted against X_{SDS}, showing that the chromophore λ_{max} is dependent on X_{SDS}. Error bars represent the standard deviation of three independent measurements.

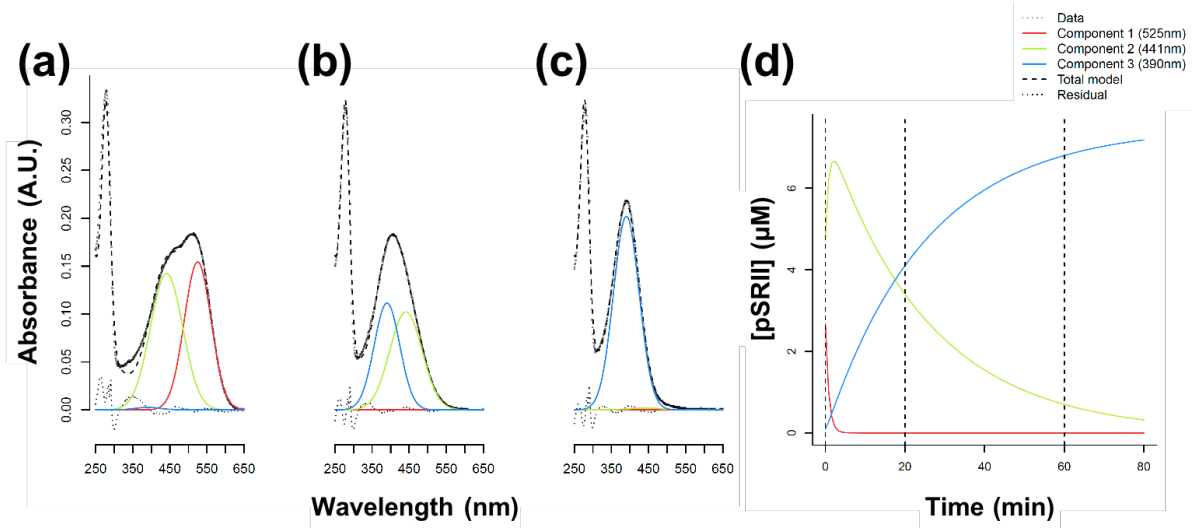


Figure 2.15: Nonlinear least squares fitting of UV/vis spectra
Time-resolved UV/vis spectra of pSRII unfolded in 0.980 X_{SDS} (3% SDS) are shown for **(a)** t = 0 min (dead-time = 25 s), **(b)** t = 20 min, and **(c)** t = 60 min. Nonlinear least squares fitting was used to deconvolute the spectra into three spectral components (red, lime and blue lines). Goodness-of-fit can be assessed from the residual and the total model taken as the sum of the three spectral components. **(d)** The concentration of each fitted component is plotted against time.

2.3.2.3 Tertiary structure changes

pSRII contains six tryptophan residues, which are mainly located in close proximity to the retinal chromophore (Figure 2.16a). Their intrinsic fluorescence is quenched due to very efficient energy transfer to the retinal by Förster resonance energy transfer (FRET) [301]. Figure 2.16b shows the emission spectra of pSRII denatured in different X_{SDS} at pH 6.0 for 2 h. Increased tryptophan fluorescence intensities were observed, as Schiff base hydrolysis leads to the removal of retinal from its binding pocket and away from the tryptophan residues. Tryptophan fluorescence of folded pSRII (0 X_{SDS}) has λ_{max} of 337 nm, which became gradually red-shifted with increasing X_{SDS} up to 343 nm in 0.998 X_{SDS} (30% SDS in 0.1% c7-DHPC). This indicates that some tryptophan residues have become more solvent-exposed upon protein unfolding. pSRII unfolded in SDS + 0.5% c7-DHPC (i.e. the conditions chosen for the kinetic studies) yielded similar observations (Figure 2.11d).

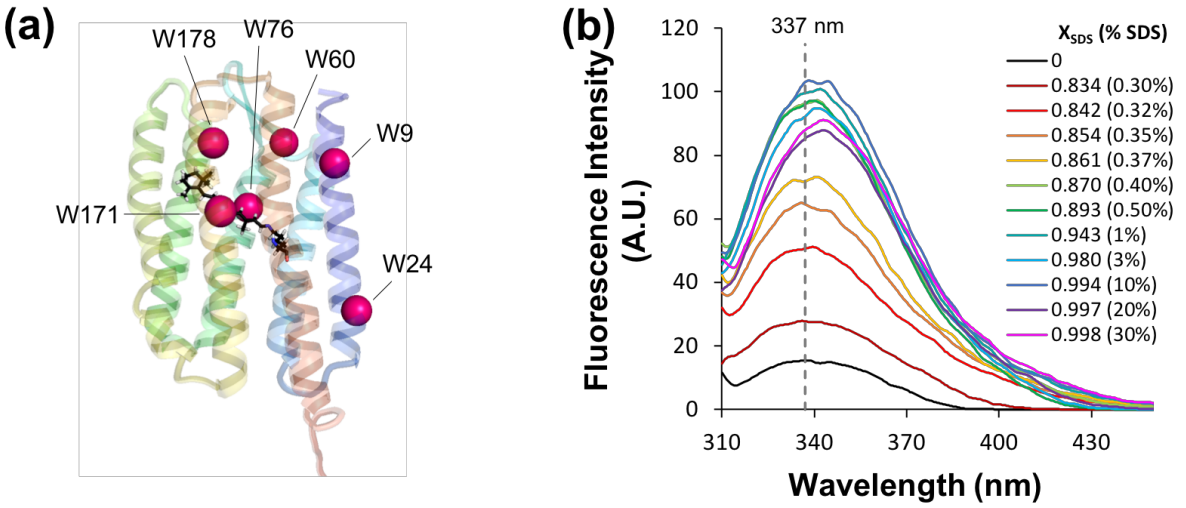


Figure 2.16: Tryptophan fluorescence of SDS-denatured pSRII at pH 6.0
(a) Solution-state NMR structure of pSRII (PDB 2KSY), with tryptophan residues shown as pink spheres. **(b)** Tryptophan fluorescence emission spectra of 7 µM pSRII unfolded for 2 h in 0–0.998 X_{SDS} (0–30% SDS in 0.1% c7-DHPC). Emission λ_{max} of pSRII in 0 X_{SDS} is shown as dotted line.

2.3.2.4 The unfolding transition

The unfolding transition of pSRII in the presence of SDS + 0.1% c7-DHPC vs. SDS + 0.5% c7-DHPC was compared using CD, UV/vis spectroscopy and tryptophan fluorescence (Figure 2.17). The unfolding transition for pSRII was ~0.81 X_{SDS} in SDS + 0.5% c7-DHPC, and ~0.86 X_{SDS} in SDS + 0.1% c7-DHPC. These values were very comparable across the

three biophysical techniques. Notably, pSRII unfolded to the same extent in SDS + 0.1% c7-DHPC vs. in SDS + 0.5% c7-DHPC, thus justifying the comparison of biophysical data obtained under these conditions.

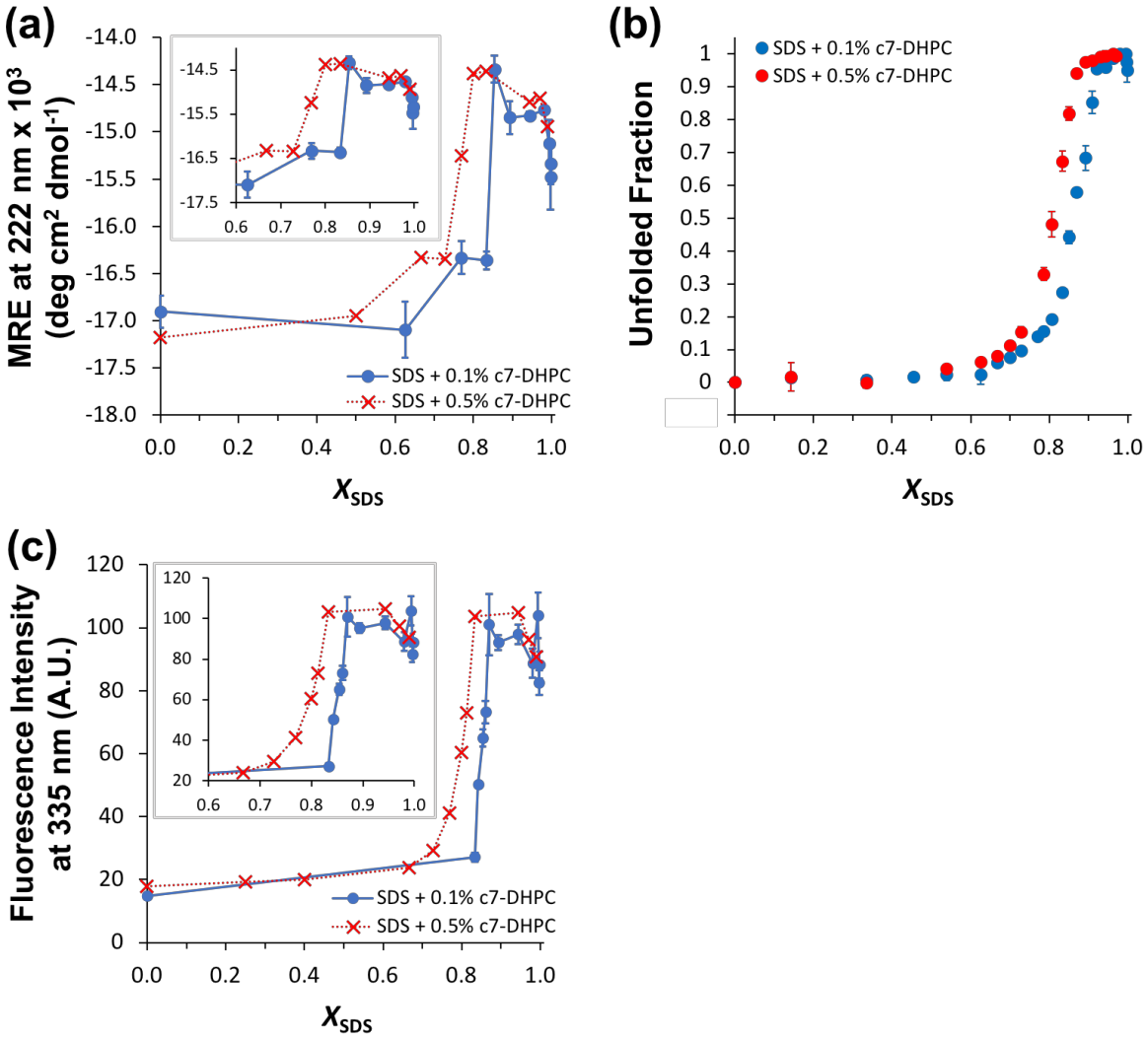


Figure 2.17: Unfolding transition of pSRII in SDS/c7-DHPC determined by (a) CD, (b) UV/vis spectroscopy, and (c) tryptophan fluorescence
Error bars represent the standard deviation of three independent measurements.

2.3.2.5 Solvent accessibility of the binding pocket

Hydroxylamine serves as a probe for monitoring protein conformational changes near the Schiff base in retinal-binding proteins [302,303]. Hydroxylamine becomes the preferred nucleophile for Schiff base cleavage instead of water, leading to cleavage of the PSB and formation of retinal oxime, which absorbs at 360 nm (Figure 2.18a) [304]. In the presence of 10 mM hydroxylamine, pSRII in 0 X_{SDS} at pH 6.0 exhibited slow retinal PSB hydrolysis, suggesting that the retinal binding pocket is relatively solvent-inaccessible (Figure 2.18b(i)). The rate of chromophore bleaching increased significantly in the presence of 0.980 X_{SDS} (3% SDS), with complete formation of retinal oxime within 30 min (Figure 2.18b(ii)). This indicates that SDS leads to an increase in solvent accessibility of the retinal binding pocket. Tryptophan fluorescence spectra of hydroxylamine-bleached pSRII, both in the absence and presence of SDS, showed an increase in fluorescence intensity and X_{SDS}-dependent red shift of the peak (Figure 2.18c), consistent with the release of retinal oxime and increase in solvent accessibility of the binding pocket. In contrary to bacteriorhodopsin and bovine rhodopsin where the bleaching of purple membranes and bovine rod outer segment disk membrane in the presence of hydroxylamine did not lead to appreciable changes in the far-UV CD spectra [305,306], the presence of hydroxylamine led to greater loss of pSRII secondary structure beyond the unfolding transition while the transition point remained the same (Figure 2.18d). This suggests that some non-native secondary structure elements of pSRII are being stabilised by SDS/c7-DHPC mixed micelle in the absence of hydroxylamine but not in the presence of hydroxylamine, possibly due to changes in detergent micelle structure and/or protein-micelle interactions in the presence of hydroxylamine.

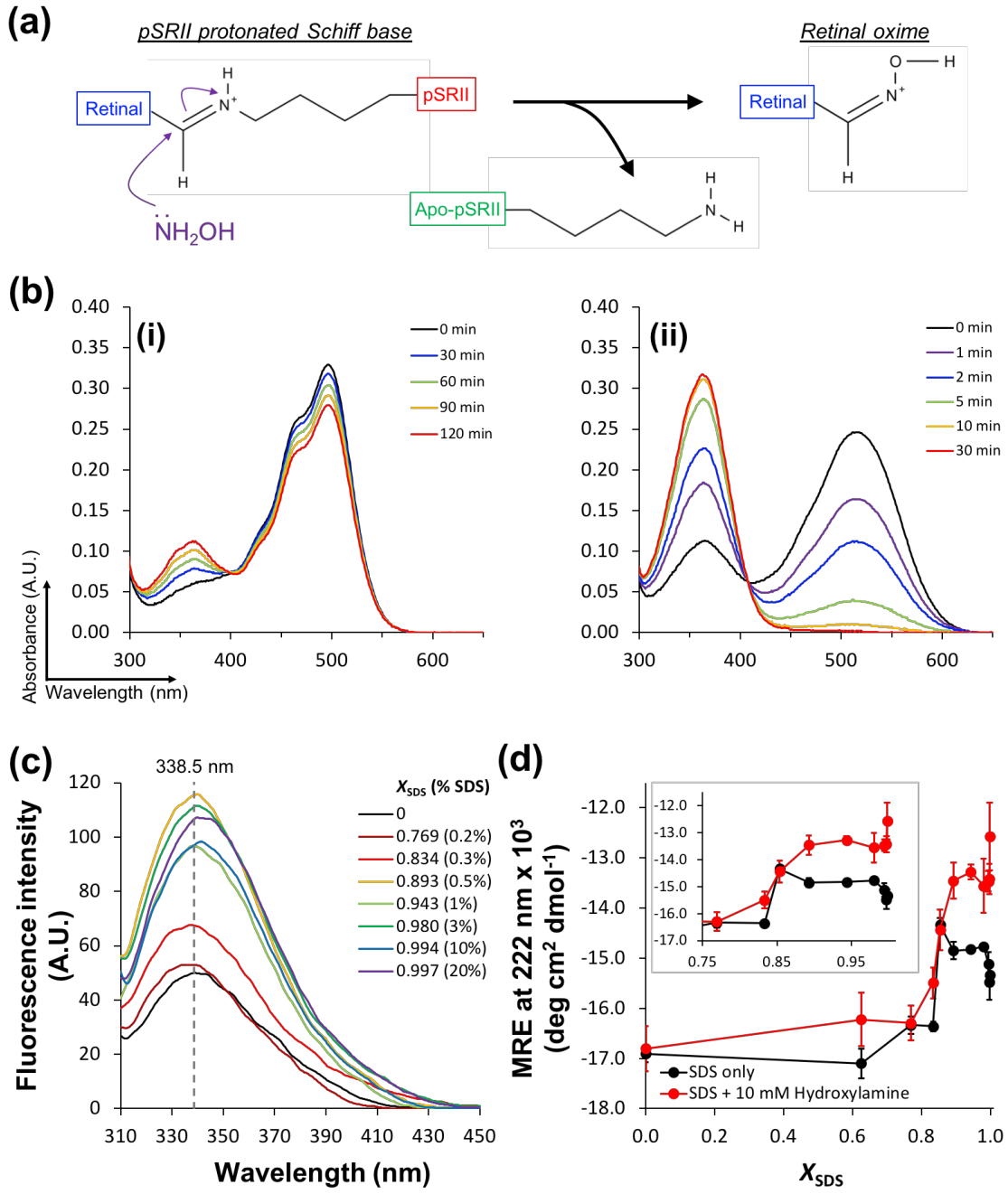


Figure 2.18: Assessing solvent accessibility of the retinal binding pocket using hydroxylamine at pH 6.0

(a) Reaction mechanism of hydroxylamine with pSRII PSB to form retinal oxime. Structures were drawn using MolView. **(b)** Time-resolved UV/vis spectra of pSRII in **(i)** 10 mM hydroxylamine (0 X_{SDS}) and **(ii)** 0.980 X_{SDS} (3% SDS) + 10 mM hydroxylamine in 0.1% c7-DHPC. **(c)** Emission spectra of 7 µM pSRII unfolded for 2 h in 0–0.998 X_{SDS} + 10 mM hydroxylamine in 0.1% c7-DHPC. Emission λ_{max} (338.5 nm) of pSRII in 0 X_{SDS} + 10 mM hydroxylamine is shown as dotted line. **(d)** MRE at 222 nm for pSRII in 0–0.998 X_{SDS} + 10 mM hydroxylamine. Inset shows expanded view for the 0.75–1.00 X_{SDS} region. Error bars represent the standard deviation of three independent measurements.

2.3.3 Kinetic measurements of pSRII unfolding

The results from biophysical studies show that pSRII unfolds in SDS at pH 6.0 to form SO_{390} via a SR_{440} intermediate, as summarised in Equation 2.20:

$$\text{pSRII} \underset{\boldsymbol{k_f}}{\overset{\boldsymbol{k_u}}{\rightleftharpoons}} \text{SR}_{440} \overset{\boldsymbol{k_h}}{\rightarrow} \text{SO}_{390} \tag{2.20}$$

where k_u, k_f and k_h are the first-order rate constants for unfolding, folding and retinal Schiff base hydrolysis, respectively.

In the SO_{390} state, most of the secondary structure remains while tertiary structure is disrupted.

Since high X_{SDS} were used for the unfolding reactions ($k_u \gg k_f$), the unfolding model could be further simplified to yield Equation 2.21:

$$\text{pSRII} \overset{k_u}{\rightarrow} \text{SR}_{440} \overset{k_h}{\rightarrow} \text{SO}_{390} \tag{2.21}$$

To gain further insights into the unfolding pathway, the kinetics of pSRII unfolding were monitored by UV/vis spectroscopy and the time courses fitted using nonlinear least squares methods (see Materials and Methods). Kinetic studies were done in the presence of excess concentrations of SDS/c7-DHPC mixed micelles (SDS + 0.5% c7-DHPC) to obtain accurate rate constants and to avoid multiple micelle occupancy.

Two rate constants, k_u and k_h, were extracted, describing changes in retinal chromophore absorbance when pSRII was unfolded in SDS at pH 6.0 (no hydroxylamine). k_u, the first-order rate constant for the formation of SR_{440}, increases with increasing X_{SDS} (Table 2.4). k_u is $(44 \pm 9) \times 10^{-4}$ s^{-1} in 0.893 X_{SDS} (2.5% SDS in 0.5% c7-DHPC), about 120-fold smaller than the unfolding rate constant of 0.54 s^{-1} for bacteriorhodopsin in 0.882 X_{SDS} in SDS/DMPC/CHAPSO micelles [116]. Bovine rhodopsin unfolds in 0.972 X_{SDS} (1% SDS in 0.05% DDM) with the time constant 0.32 ± 0.01 s [135], corresponding to k_u of 3.125 s^{-1}, which is about 700-fold larger than that for the unfolding of pSRII. Plotting the natural logarithm of k_u against X_{SDS} yields the unfolding arm of the chevron plot (see Section 2.3.4), where ln k_u increases linearly against X_{SDS}. Minor non-linearity observed in high X_{SDS} can be attributed to the formation of cylindrical micelles at high SDS concentrations [307,308].

k_h, the rate constant of Schiff base hydrolysis, is independent of SDS concentration (Table 2.4). k_h is (4.2 ± 0.6) × 10^{-4} s^{-1} in 0.834–0.971 X_{SDS} (1.5–10% SDS in 0.5% c7-DHPC), similar to but slightly smaller than ~9×10^{-4} s^{-1}, the rate constant for free retinal formation from the 440-nm state of bacteriorhodopsin (bR_{440}) [116]. This suggests that SR_{440} and bR_{440} could have similar solvent accessibility to their binding pockets, but k_h might also be rate-limited by other factors such as the reactivity of the Schiff base, possibly due to different unfolding conditions being used in these studies.

X_{SDS}	Rate constants (×10^{-4} s^{-1})	
	k_u	k_h
0.834	18 ± 3	3.7 ± 0.2
0.870	36 ± 11	3.9 ± 0.3
0.893	44 ± 9	4.5 ± 0.3
0.909	62 ± 12	3.8 ± 0.4
0.930	73 ± 4	4.1 ± 0.9
0.943	106 ± 8	4.2 ± 0.9
0.962	98 ± 17	4.5 ± 0.8
0.971	111 ± 24	4.4 ± 0.6

Table 2.4: Rate constants for pSRII unfolding in SDS/c7-DHPC mixed micelles
k_u is the first-order rate constant for the formation of SR_{440}, and k_h is the rate constant for Schiff base hydrolysis. Errors are reported as standard deviations of three independent measurements.

To compare results across different conditions and biophysical techniques, changes over time of MRE at 222 nm, retinal UV/vis absorbance, and tryptophan fluorescence emission at 335 nm were recorded for pSRII unfolded in different X_{SDS} in 0.1% c7-DHPC at pH 6.0, both in the absence and presence of 10 mM hydroxylamine. The time courses were fitted to the appropriate models using nonlinear least squares fitting (see Materials and Methods, Figures 2.15 and 2.19) and the rate constants were extracted (Figure 2.20).

These rate constants differ slightly from those for pSRII unfolding in SDS + 0.5% c7-DHPC (see Table 2.4) due to possible complications with multiple micelle occupancy under low c7-DHPC concentrations. Nevertheless, good consistency was obtained across the three

techniques, indicating that rate constants extracted by UV/vis spectroscopy for pSRII unfolding in SDS + 0.5% c7-DHPC would have been comparable to those determined by CD and tryptophan fluorescence.

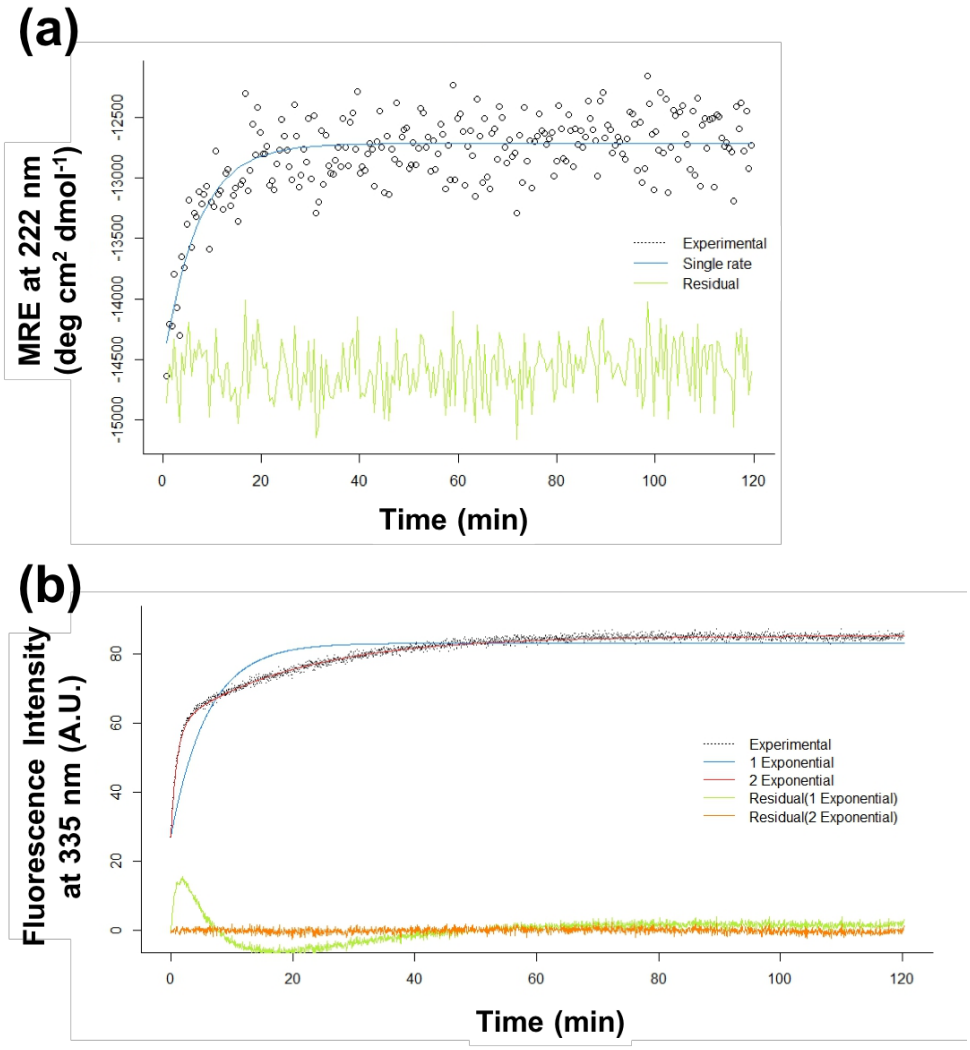


Figure 2.19: Examples of nonlinear least squares fitting for time-courses recorded by (a) far-UV CD, and (b) tryptophan fluorescence.
(a) Far-UV CD timecourse data (white circles) were adequately fitted with a single exponential model (blue line), as judged by the residual (green line). **(b)** Tryptophan fluorescence timecourse data (black dots) could not be adequately fitted with a single exponential (blue line), as judged by the non-random residual trace (green line), while a double exponential model (red line) fits the data adequately, as determined from the corresponding residual (orange line).

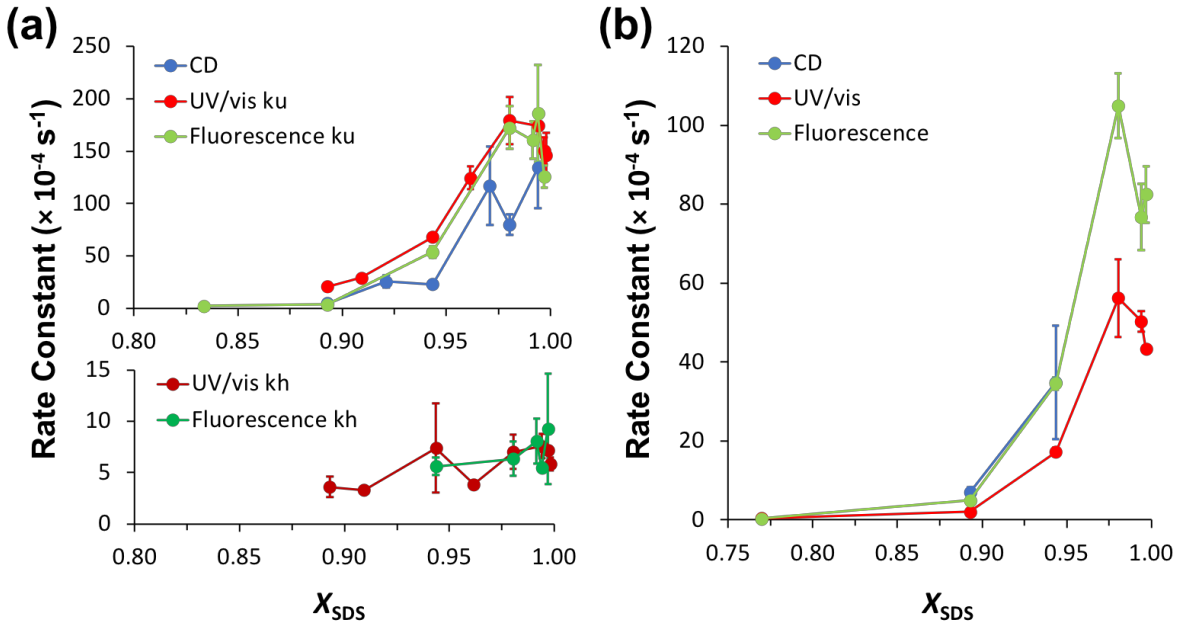


Figure 2.20: X_{SDS}-dependence of the rate constants for pSRII unfolding
(a) In the presence of SDS only (no hydroxylamine), two rate constants, k_u and k_h, were extracted to describe protein unfolding and retinal Schiff base hydrolysis, respectively. **(b)** In the presence of 10 mM hydroxylamine, only one rate constant was extracted. Error bars in both panels represent the standard deviation of three independent measurements.

Secondary structure changes occur as a one-step process, as evidenced by good monoexponential fits of changes in MRE at 222 nm (Figure 2.19a). Similar rate constants were obtained regardless of the absence or presence of hydroxylamine (Figure 2.20). Tertiary structure changes, monitored by UV/vis spectroscopy and tryptophan fluorescence time-courses, occur in two steps: protein unfolding to form SR_{440} and Schiff base hydrolysis to yield SO_{390}, respectively. The rate constants extracted from tryptophan fluorescence time-courses were similar to those from UV/vis time-courses for most X_{SDS}, hence confirming that the increase in fluorescence intensity during unfolding was primarily due to un-quenching of tryptophan fluorescence while retinal was being removed from its binding pocket (Figure 2.20). At 0.893 X_{SDS} (0.5% SDS + 0.1% c7-DHPC) and 0.943 X_{SDS} (1% SDS + 0.1% c7-DHPC), the unfolding rate constant measured by tryptophan fluorescence was smaller than that measured by UV/vis spectroscopy, suggesting that fluorescence remains partially quenched for one or more tryptophan residues near the partially loosened retinal binding pocket. Schiff base hydrolysis was fast in the presence of hydroxylamine and SDS at pH 6.0, such that retinal oxime formed rapidly without accumulation of the SR_{440}

intermediate, and only one rate constant was extracted across all biophysical techniques (Figure 2.20b).

In summary, exposure of pSRII to SDS led to X_{SDS}-dependent protein unfolding followed by X_{SDS}-independent retinal Schiff base hydrolysis. The rate limiting step depends on the relative rates of these two processes, and determines the rate constant(s) that can be extracted from the time-courses.

2.3.4 Refolding from SR_{440}

Reversibility of *in vitro* unfolding is important for determining whether the denatured states are on-pathway species. SDS-denatured pSRII constitutes the starting point of *in vitro* refolding experiments. The reversibility of unfolding and the kinetics of refolding reactions were characterised in order to understand the folding pathway of pSRII. As there are two stages (protein unfolding and retinal Schiff base hydrolysis) involved in obtaining SDS-denatured apo-sensory opsin, the reversibility of pSRII unfolding was examined with refolding reactions initiated from SDS-denatured pSRII with covalently-attached retinal (SR_{440}) or from SDS-denatured apo-sensory opsin (SO_{390}). Successful refolding was assessed by the recovery of the native chromophore absorbance at 498 nm with characteristic vibronic side bands at 460, 420 and 370 nm.

SR_{440} was first attained by incubating pSRII in 0.943 X_{SDS} (5% SDS + 0.5% c7-DHPC) for 4 min. Refolding was then initiated by 12-fold dilution of SR_{440} into refolding buffer containing different X_{SDS} and monitored by UV/vis spectroscopy. All refolding kinetics were measured in the presence of 0.5% c7-DHPC to minimise the effects of multiple micelle occupancy. Refolding buffer included 0.500–0.700 X_{SDS} (0.30–0.70% SDS) for measuring refolding rates in 0.698–0.779 X_{SDS} (0.69–1.06% SDS).

Absorption spectra were taken at fixed time points during refolding, and deconvoluted by nonlinear least squares fitting in order to assess the rate at which the native chromophore is being reformed (Table 2.5).

Figure 2.21a shows the spectra recorded in 0.698 X_{SDS}, where the rate constant of folding (k_f) was significantly greater than the rate constant of retinal hydrolysis (k_h), and in 0.779 X_{SDS}, where k_f was similar to k_h. In both cases, as the reaction progressed, absorbance at 440 nm decreased, while absorbance at 390 nm and ~500 nm increased simultaneously, albeit at different rates. This indicates that formation of folded pSRII is fast, whereas formation of SO_{390}, either directly from SR_{440} or from slow unfolding of refolded pSRII, is slow and observable within 2 h only in ≥ 0.745 X_{SDS}.

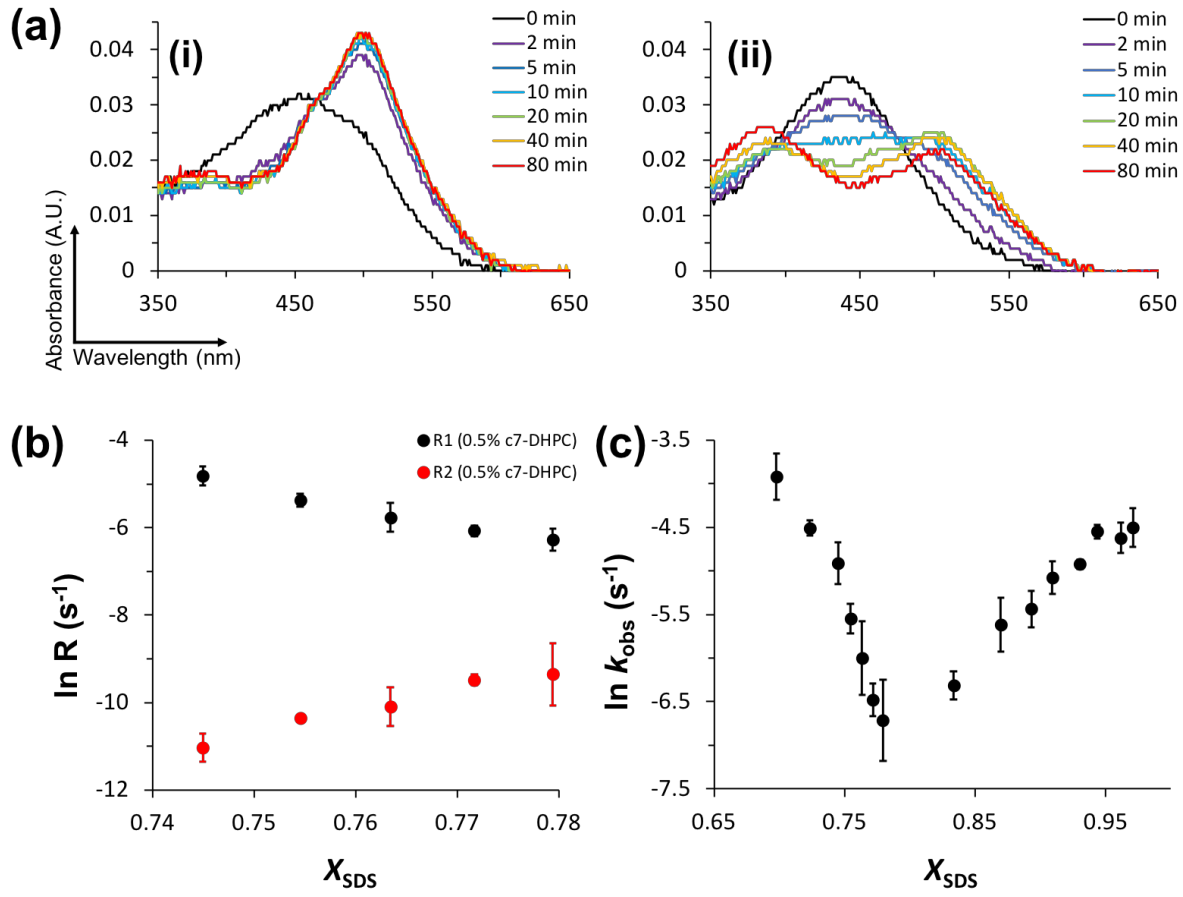


Figure 2.21: Refolding of pSRII from the unfolded state (SR_{440})
(a) Absorption spectra for refolding from SR_{440} in **(i)** 0.698 X_{SDS} and **(ii)** 0.779 X_{SDS}, both in 0.5% c7-DHPC. Both regeneration of native chromophore (498 nm) and formation of SO_{390} can be observed, with the rate of refolding being dependent on X_{SDS}. **(b)** Changes in absorbance at ~500 nm in high X_{SDS} can be fitted to a double exponential to yield two macroscopic rate constants, R_1 and R_2. **(c)** Chevron plot summarising the extracted microscopic unfolding and refolding rates, k_u and k_f. Error bars in panels b and c represent the standard deviation of three independent measurements.

X_{SDS}	Rate constants × 10^{-4} (s^{-1})			
	R_1	R_2	k_f	k_u
0.698	-	-	199 ± 52	-
0.723	-	-	110 ± 10	-
0.745	81 ± 18	0.16 ± 0.05	74 ± 18	3.5 ± 1.4
0.755	46 ± 7	0.32 ± 0.01	39 ± 7	4.0 ± 0.6
0.763	31 ± 10	0.41 ± 0.18	25 ± 10	3.3 ± 1.8
0.772	23 ± 3	0.76 ± 0.09	15 ± 3	4.7 ± 0.8
0.779	19 ± 5	0.87 ± 0.62	12 ± 6	3.9 ± 2.9

Table 2.5: Rate constants extracted for refolding pSRII from SR_{440}
Under conditions where the rate of retinal hydrolysis is very slow compared to the rate of folding (0.698–0.723 X_{SDS}), changes in the fitted absorbance at ~500 nm over time were approximated to a monoexponential model (Equation 2.22). In higher X_{SDS}, changes in the fitted absorbance at ~500 nm over time were fitted to a double exponential model to yield the macroscopic rate constants R_1 and R_2 (Equation 2.23). The microscopic rate constants k_f and k_u were extracted using experimentally determined values of R_1, R_2 and k_h (Equations 2.24–2.26). Errors are reported as standard deviations of three independent measurements.

For ≤ 0.723 X_{SDS}, where the rate of retinal hydrolysis is very slow compared to the rate of folding ($k_h \ll k_f$), the time dependence of pSRII recovery was approximated with a single exponential model (Figure 2.22a–c):

$$A_{500} \sim A_1\left(1 - e^{-k_f t}\right) + c \tag{2.22}$$

where the observed rate constant was approximated as the *de novo* refolding rate k_f.

For ≥ 0.745 X_{SDS}, a double exponential model would be required (Figure 2.22d–f):

$$A_{500} \sim A_1(1 - e^{-R_1 t}) + A_2(e^{-R_2 t}) + c \tag{2.23}$$

where A_1 and A_2 are the amplitudes of the fast phase and the slow phase, respectively, and R_1 and R_2, the apparent rate constants (Figure 2.21b), are the solutions to the following quadratic equation [118]:

$$R^2 - \left(k_f + k_u + k_h\right)R + k_u k_h = 0 \tag{2.24}$$

k_f and k_u were calculated using the following relationships and experimentally determined values of R_1, R_2 and k_h (Table 2.5):

$$R_1 R_2 = k_u k_h \tag{2.25}$$

$$R_1 + R_2 = k_f + k_u + k_h \tag{2.26}$$

Based on experimental unfolding kinetic measurements, k_h is independent of X_{SDS} (Table 2.4), and is taken as $(3.72 \pm 0.17) \times 10^{-4}$ s^{-1} for 0.834 X_{SDS} in 0.5% c7-DHPC. It is noted that values of k_u obtained from refolding kinetic measurements tend to be unreliable, as they are dominated by errors in k_f and k_h.

In contrast to the unfolding arm of the chevron plot (ln k_f and ln k_u against X_{SDS}), which is essentially linear, the folding arm showed clear deviation from a linear trend ('roll-over') (Figure 2.21c). This result strongly indicates that the formation of tertiary structure packing during refolding of SR_{440} is more complex than the two-state kinetics that has been observed for the equivalent process in bacteriorhodopsin [116,309].

For a protein that has a tendency to oligomerise, non-linearities in chevron plots could be attributed to multiple micelle occupancy, transient aggregates or kinetic intermediates [6]. To determine whether non-linearity in the folding arm is due to any of these factors, unfolding and refolding time-courses were recorded in increased detergent-to-protein ratios [310], obtained by increasing the detergent concentration and/or decreasing the protein concentration. The unfolding and refolding rate constants obtained (Figure 2.23a–b) were superimposable with those in Figure 2.21c. SDS-PAGE on samples taken at various time points up to 2 h of refolding in different X_{SDS} also showed no increase in the amount of stably-formed oligomers compared to folded pSRII (Figure 2.23c). These results indicate that there is no evidence for aggregation or multiple micelle occupancy being the origins of non-linearity in the chevron plot, suggesting a more complex refolding mechanism for pSRII.

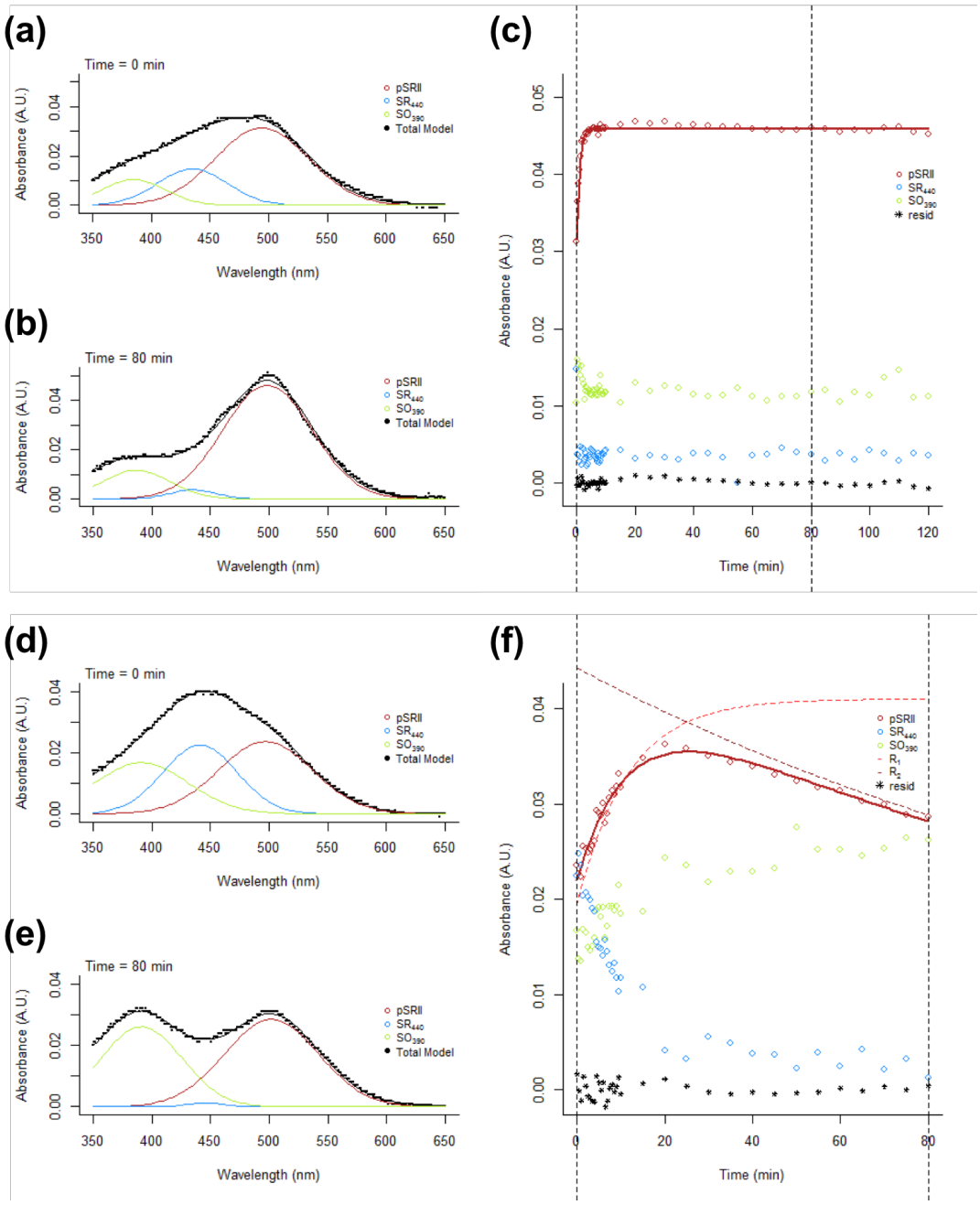


Figure 2.22: Examples of nonlinear least squares fitting for UV/vis time-courses of refolding from SR_{440} in 0.723 X_{SDS} (a–c) and 0.806 X_{SDS} (d–f)
Time-resolved UV/vis spectra of refolding from SR_{440} are shown for t = 0 min (dead-time = 20 s) **(a, d)**, and t = 80 min **(b, f)**. The spectra were deconvoluted by nonlinear least squares (NLS) fitting to yield three spectral components (red, lime and blue lines). Goodness-of-fit can be assessed from the total model taken as sum of the three spectral components. The absorbance of each fitted spectral component is plotted in panels c and f. Changes in absorbance at ~500 nm during refolding in 0.723 X_{SDS} was fitted to a mono-exponential model to yield one rate constant for the recovery of the native chromophore. A double exponential model was used to describe refolding in 0.806 X_{SDS} to yield two rate constants (R_1 and R_2) for the fast recovery of the native chromophore and slow decay to form SO_{390}, respectively.

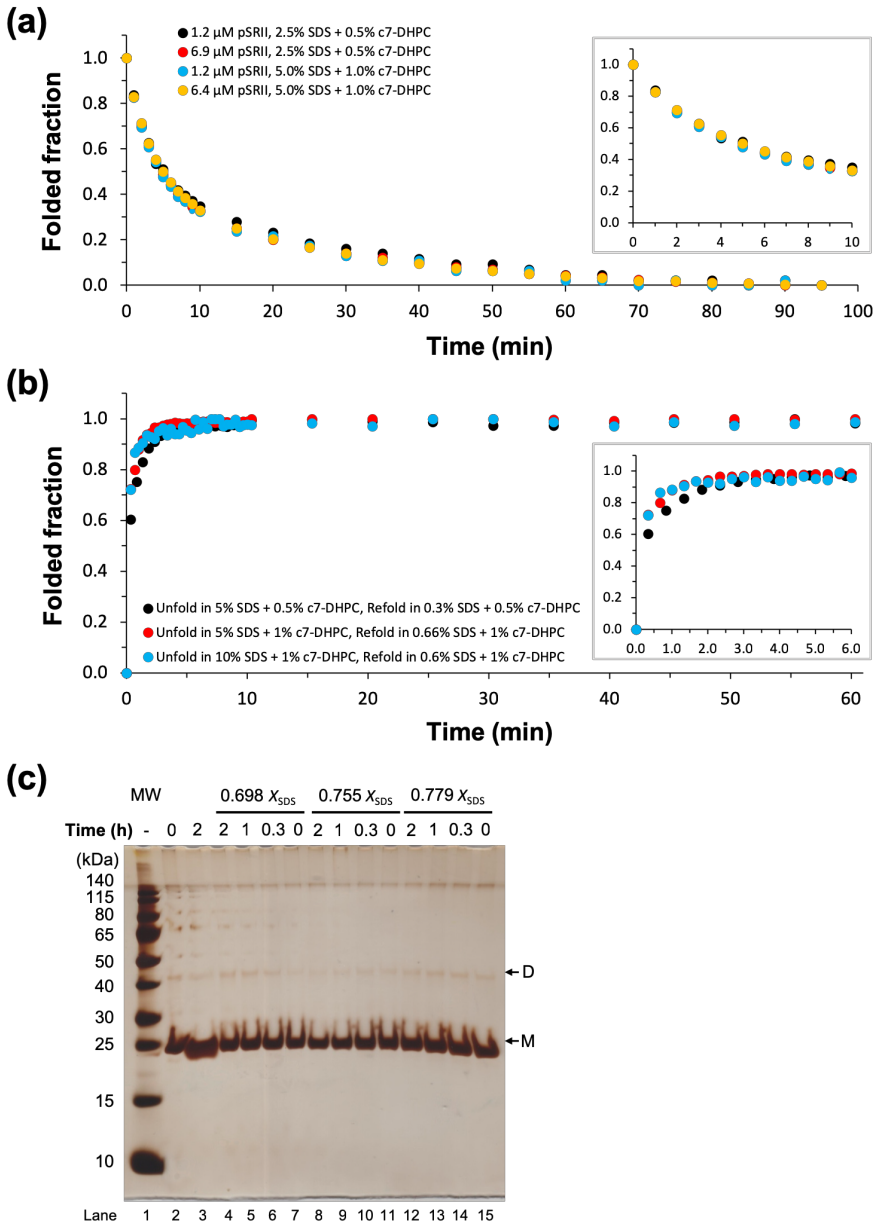


Figure 2.23: SR_{440} refolding kinetics were not influenced by aggregates or multiple micelle occupancy
(a) Normalised changes in A_{498} for pSRII unfolding in 0.893 X_{SDS} and **(b)** nonlinear least squares fitted A_{500} for SR_{440} refolding in 0.698 X_{SDS}, both showing that unfolding and refolding rates are not dependent on protein concentration or detergent concentration. 1.2 µM pSRII was used in refolding experiments shown in panel b. Insets in both panels show expanded views of the earliest time points. **(c)** SDS-PAGE for refolding from SR_{440} for different durations (0 min, 20 min, 1 h, 2 h) in the presence of different X_{SDS} in 0.5% c7-DHPC, showing no evidence for the formation of stable aggregates. Lane 1, molecular weight marker; lanes 2–3, pSRII unfolded in 0.893 X_{SDS} for 0 min and 2 h, respectively; lanes 4–15, pSRII was first unfolded for 4 min in 0.893 X_{SDS} to obtain SR_{440}, followed by refolding for 0 min, 20 min, 1 h and 2 h in 0.698 X_{SDS} (lanes 4–7), 0.755 X_{SDS} (lanes 8–11) and 0.779 X_{SDS} (lanes 12–15). The arrows indicate the expected migration of monomeric (M) and dimeric (D) pSRII.

2.3.5 Refolding from SO_{390}

Refolding from the SO_{390} apo-protein state was explored to determine whether retinal Schiff base hydrolysis is reversible. For retinal-binding proteins, formation of the retinal Schiff base linkage and correct protein folding (see Section 2.3.4 on refolding from SR_{440}) are both crucial for formation of the native chromophore in a functional binding pocket.

SO_{390} was first attained by incubating pSRII in 0.893 X_{SDS} (2.5% SDS in 0.5% c7-DHPC) for 2 h at room temperature. Retinal was not removed after unfolding and no additional retinal was introduced. Refolding was initiated by 12-fold dilution of SO_{390} into refolding buffer and monitored for 2 h by UV/vis spectroscopy. The refolding buffer included 0–0.294 X_{SDS} (0–0.125% SDS in 0.5% c7-DHPC) to enable measurement of refolding kinetics in 0.410–0.519 X_{SDS} (0.208–0.323% SDS in 0.5% c7-DHPC).

Time-courses of refolding from SO_{390} showed an isosbestic point at 420 nm (Figure 2.24a), suggesting that the rate-limiting step of the reaction is the reformation of retinal Schiff base, assuming that the reaction proceeds through SR_{440} (i.e. the same pathway as unfolding).

The time-courses were modelled to Equation 2.27 as pseudo first-order reactions:

$$SO_{390} \xrightarrow{k_{app}} \text{pSRII} \tag{2.27}$$

where k_{app} is the apparent rate for the reconstitution of the native chromophore, taking into account both retinal re-attachment (reformation of the Schiff base) and protein folding.

k_{app} in 0.410–0.519 X_{SDS} ranged from 1.7×10^{-4} to 12.9×10^{-4} s^{-1} (Table 2.6), similar to the rate constant of 17×10^{-4} s^{-1} for the refolding of bacteriorhodopsin from bO_{390} in 0.472 X_{SDS} [116]. This suggests that the rate of refolding from SO_{390} is limited by Schiff base reformation. The natural logarithm of k_{app} ($\ln k_{app}$) appeared to be linearly-dependent against X_{SDS} in 0.410–0.519 X_{SDS} (Figure 2.24b), though the fast k_{app} in < 0.4 X_{SDS} precludes its measurement by means of manual mixing.

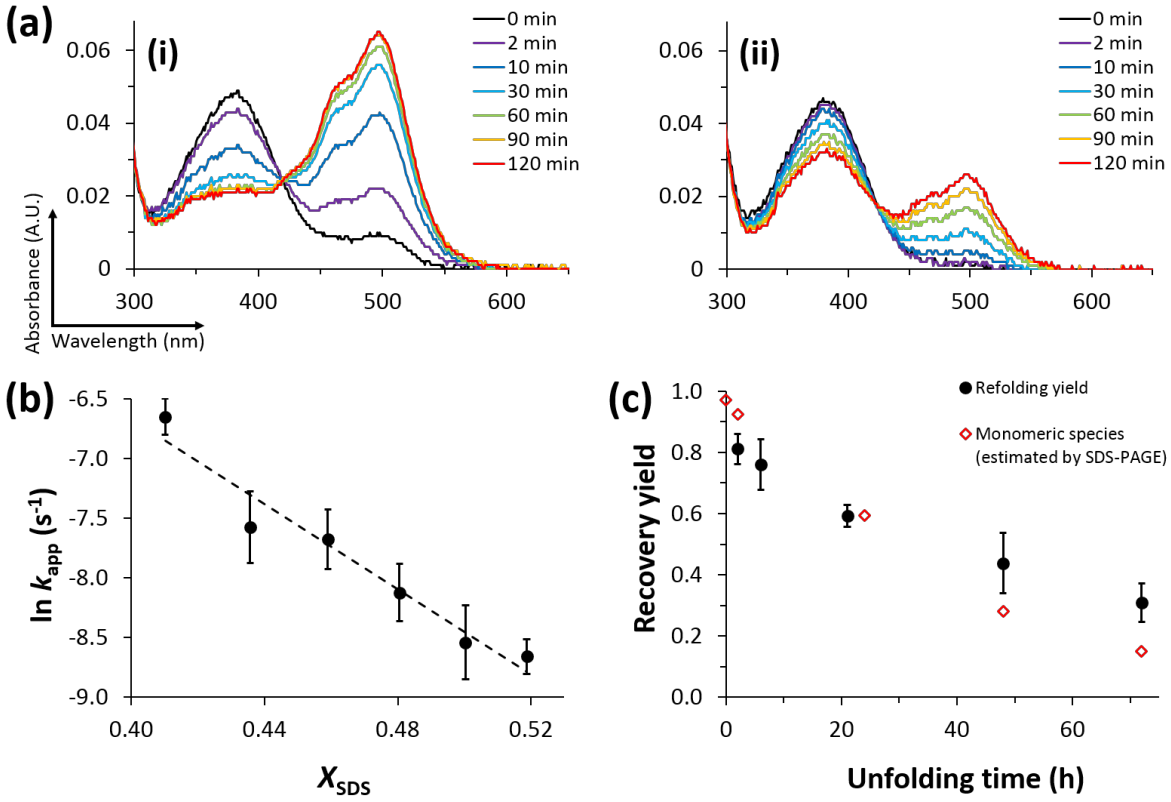


Figure 2.24: Refolding of pSRII from the apo-protein state (SO_{390})
(a) Absorption spectra for refolding from SO_{390} in **(i)** 0.410 X_{SDS} and **(ii)** 0.519 X_{SDS}, both in 0.5% c7-DHPC. **(b)** Natural logarithm of the apparent refolding rate (ln k_{app}) from SO_{390} is linearly dependent on X_{SDS}. **(c)** Refolding yield of SO_{390} decreases with increasing amount of time spent in the unfolded state, suggesting that only the monomeric species might be capable of rebinding retinal and regenerating the native chromophore. Error bars in panels b and c represent the standard deviation of three independent measurements. Results from one representative gel is shown for estimating the monomeric content by SDS-PAGE (see Figure 2.11b: lanes 6–10).

X_{SDS}	$k_{app} \times 10^{-4}$ (s^{-1})
0.410	12.9 ± 2.0
0.436	5.1 ± 1.5
0.459	4.6 ± 1.2
0.480	3.0 ± 0.7
0.500	2.0 ± 0.6
0.519	1.7 ± 0.2

Table 2.6: Apparent rate constants (k_{app}) for refolding pSRII from SO_{390}.
k_{app} takes into account both retinal re-attachment (reformation of the Schiff base) and protein folding. Errors are reported as standard deviations of three independent measurements.

Overnight incubation of the refolding reaction at room temperature led to a substantial increase in the amount of refolded pSRII compared to the recovery yield obtained after 2 h (Figure 2.25a). This additional slow phase for native chromophore recovery might suggest that (1) there is a small population of misfolded SO_{390} which requires some structural rearrangements to be capable of rebinding retinal and thus refolds at a much slower timescale, and/or (2) retinal has diffused out of a subpopulation of detergent micelles and the regeneration of SO_{390} apo-protein molecules which are embedded in such micelles would depend on the resolubilisation of retinal into these micelles.

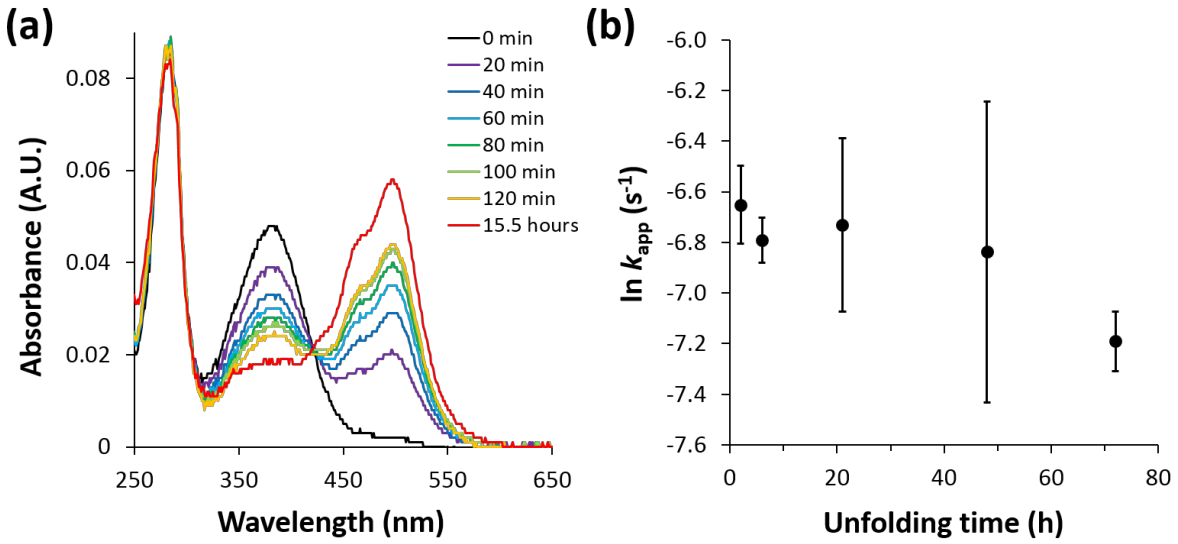


Figure 2.25: Additional recovery of the native chromophore after overnight refolding of SO_{390}
(a) Example time-course for the refolding of SO_{390} in 0.480 X_{SDS} in 0.5% c7-DHPC, showing additional recovery of the native chromophore (red trace) after refolding at room temperature overnight. **(b)** The apparent rate constant for refolding remains the same regardless of the time spent in the SO_{390} state. Error bars represent the standard deviation of three independent measurements.

Refolding yields were estimated by calculating the concentration of free retinal at the projected plateau of the single exponential fitted for the decrease in SO_{390} over the two-hour refolding time-course (Figure 2.26). Refolding yields decreased with increasing amount of time (2–72 h) during which pSRII remained unfolded in 0.893 X_{SDS}, and correlated well with the monomeric population left in the unfolded reaction assessed by SDS-PAGE (Figure 2.11b: lanes 6–10) and quantified using ImageJ (Figure 2.24c). Hence, only monomeric SO_{390} was capable of refolding to yield the native chromophore, while the oligomeric species were

irrecoverable misfolded off-pathway species. The apparent refolding rate constant was independent of the time spent in the SO_{390} state and the extent of aggregation (Figure 2.25b).

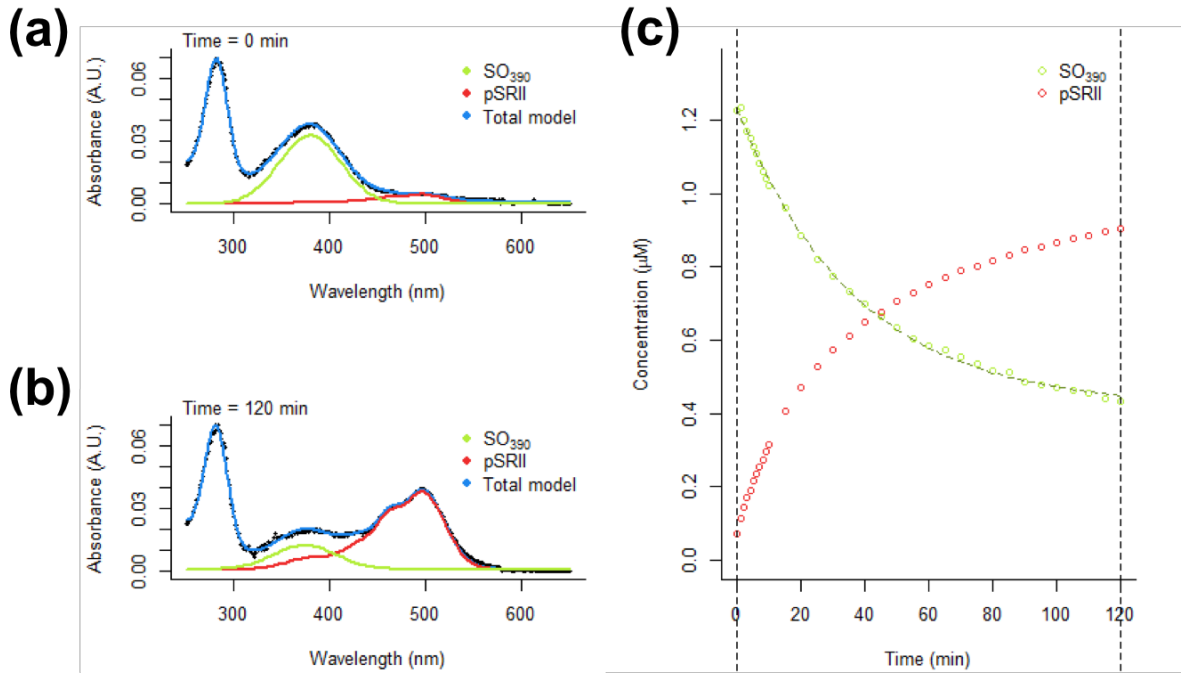


Figure 2.26: Examples of nonlinear least squares fitting of UV/vis time-courses for the refolding of SO_{390} in 0.745 X_{SDS} (a–c) and 0.806 X_{SDS} (d–f)
Time-resolved UV/vis spectra of refolding from SO_{390} are shown for t = 0 min (dead-time = 20 s) **(a)**, and t = 120 min **(b)**. The spectra were deconvoluted by nonlinear least squares fitting to yield two spectral components (red and lime lines) for the retinal chromophore. Goodness-of-fit can be assessed from the total model taken as the sum of the fitted spectral components. The absorbance of each fitted spectral component is plotted in panel c. Changes in absorbance at ~390 nm during refolding was fitted to a monoexponential model to yield one apparent rate constant for the decay of SO_{390}.

2.3.6 Characterisation of denatured states: SDS at pH 2.0

2.3.6.1 Secondary structure changes

pSRII unfolded in SDS at pH 2.0 also showed little loss in α-helical content (Figure 2.27). Essentially no difference in MRE at 222 nm was observed even after exposing pSRII to acidic pH for 19 h, suggesting slow unfolding in the absence of SDS (Figure 2.27a(i)). 0.998 X_{SDS} (30% SDS in 0.1% c7-DHPC) at pH 2.0 led to fast loss of α-helicity within seconds (Figure 2.27a(ii)), contrasting against pSRII unfolding in SDS at pH 6.0, where the loss of secondary structure occurs on timescales ranging from seconds to hours, with the rate constant being dependent on X_{SDS} (Figure 2.20). Further loss of α-helicity beyond 20 h in

SDS at pH 2.0 is attributed to aggregation, as evidenced by SDS-PAGE (Figure 2.10) and the appearance of a local minimum at 218 nm in the CD spectrum, which is characteristic of an increase in β-sheet content (Figure 2.27a(ii)). Changes in different secondary structure elements at pH 2.0 were calculated using CDPro software (Table 2.7 and Figure 2.28). Like at pH 6.0 (Table 2.3 and Figures 2.12–2.13), the fraction of regular helix (H(r)) showed significant decrease in > 0.834 X_{SDS} at pH 2.0. While the fraction of distorted helix (H(d)) and regular strand (S(r)) did not show significant changes, the fraction of distorted strand (S(d)), turns (T) and unordered (U) structures showed slight increases.

X_{SDS}	H(r)	H(d)	S(r)	S(d)	T	U
0	46.0 ± 1.4%	17.3 ± 0.7%	1.1 ± 0.2%	3.1 ± 0.4%	11.5 ± 0.6%	21.0 ± 1.2%
0.769	47.6 ± 1.0%	18.6 ± 0.4%	0.8 ± 0.4%	2.8 ± 0.3%	10.9 ± 0.4%	19.3 ± 0.3%
0.893	42.2 ± 1.3%	18.6 ± 0.6%	1.7 ± 0.4%	3.4 ± 0.2%	12.7 ± 0.3%	21.4 ± 0.5%
0.980	37.8 ± 3.3%	17.7 ± 0.2%	3.0 ± 1.8%	4.5 ± 0.2%	13.7 ± 0.9%	23.2 ± 0.7%
0.994	39.6 ± 0.6%	18.3 ± 0.3%	1.9 ± 0.4%	4.1 ± 0.2%	13.4 ± 0.1%	22.7 ± 0.1%
0.998	41.2 ± 0.7%	19.5 ± 1.6%	1.4 ± 0.5%	3.5 ± 0.8%	11.3 ± 2.4%	23.0 ± 0.9%

Table 2.7: Deconvolution of CD spectra for pSRII unfolded for 30 min in SDS at pH 2.0
Deconvolution was performed using CDPro with CONTINLL algorithm and basis set no. 10 (contains soluble and membrane proteins). H(r), regular helix; H(d), distorted helix; S(r), regular strand; S(d), distorted strand; T, turns; U, unordered. Errors are reported as standard deviations of three independent measurements.

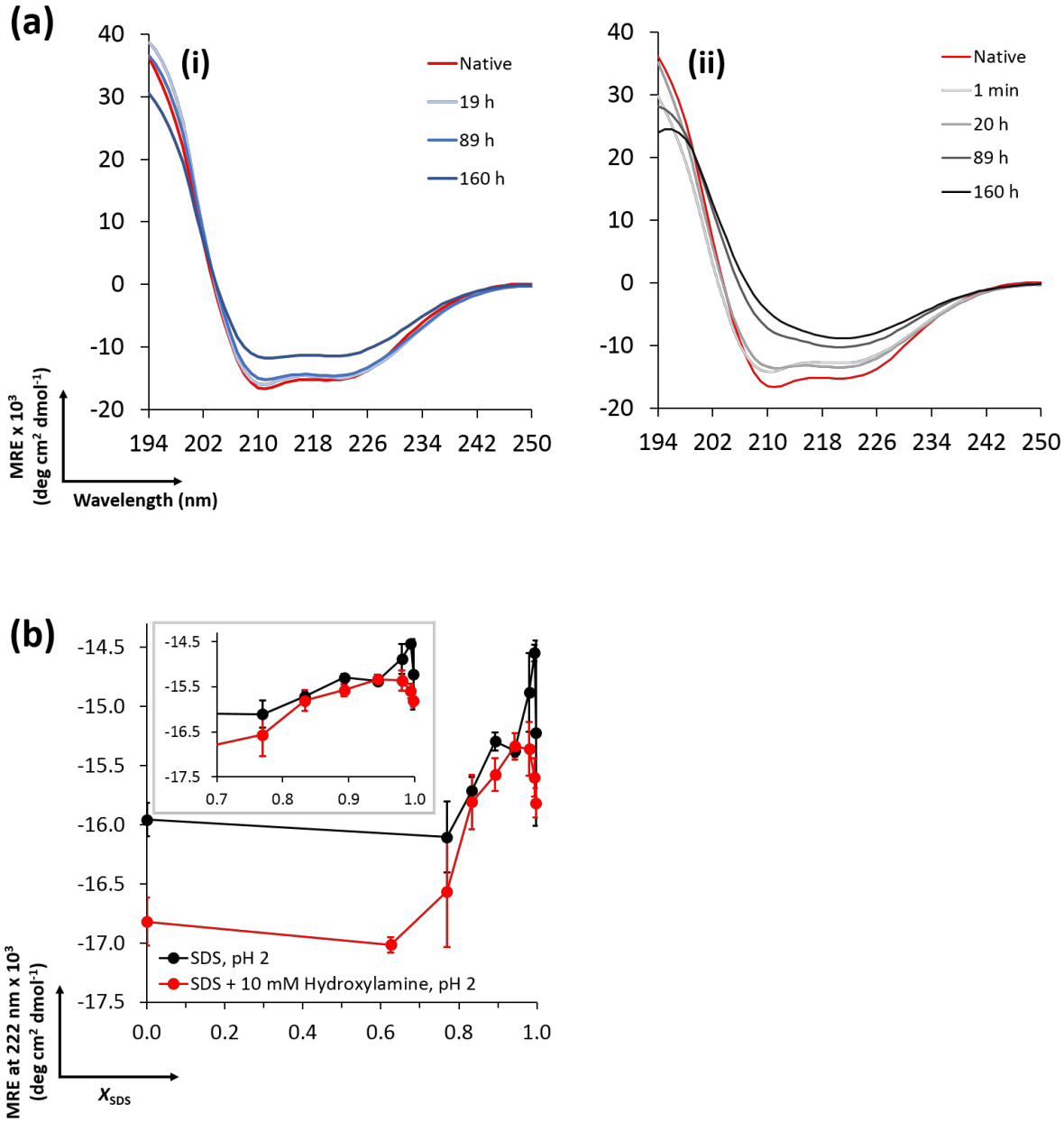


Figure 2.27: Circular dichroism of pSRII in SDS + 0.1% c7-DHPC at pH 2.0
(a) CD spectra of 7 µM (0.2 mg/ml) native pSRII and pSRII unfolded in **(i)** 0 X_{SDS} and **(ii)** 0.998 X_{SDS} (30% SDS) at pH 2.0 for up to 160 h. **(b)** MRE at 222 nm for pSRII in 0–0.998 X_{SDS} (0–30% SDS) at pH 2.0 in the absence (black trace) and presence (red trace) of 10 mM hydroxylamine. Inset shows expanded view of the 0.7–1.0 X_{SDS} region. Error bars represent the standard deviation of three independent measurements.

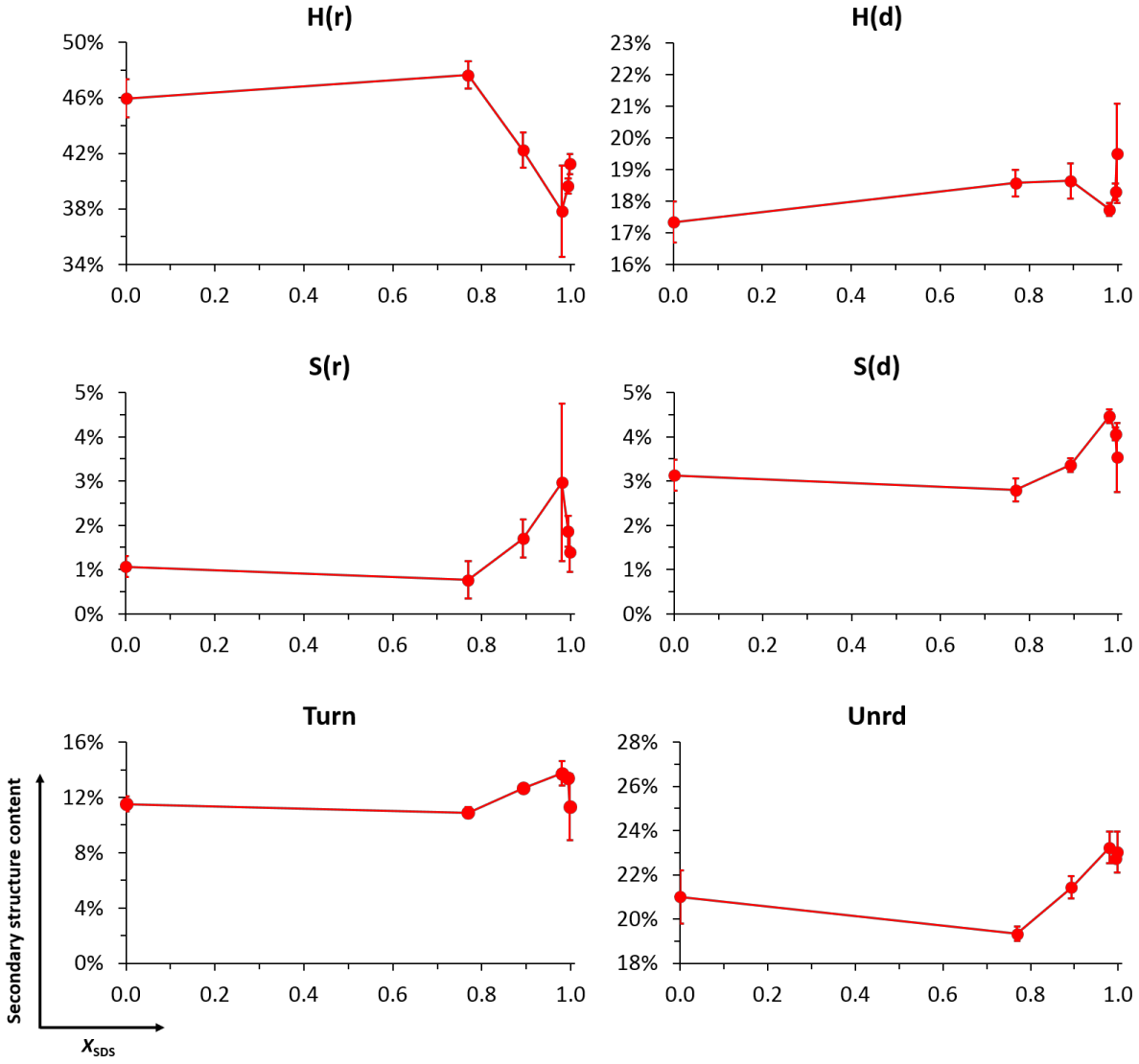


Figure 2.28: Changes in secondary structure composition of pSRII in SDS at pH 2.0
CD spectra were deconvoluted using CDPro with CONTINLL algorithm and basis set no. 10 (contains soluble and membrane proteins). H(r), regular helix; H(d), distorted helix; S(r), regular strand; S(d), distorted strand; T, turns; U, unordered. Errors are reported as standard deviations of three independent measurements.

2.3.6.2 Changes in the retinal binding pocket

Acid denaturation of pSRII at pH 2.0 (without SDS) led to very slow protein unfolding and Schiff base hydrolysis, with the reaction taking between 5 and 24 h to reach completion (Figure 2.29a). pSRII in ≥ 0.625 X_{SDS} (0.1% SDS in 0.1% c7-DHPC) at pH 2.0 led to the formation of SR_{440} within the dead-time of 25 s, followed by very slow release of free retinal over a period of more than 6 days (Figure 2.29b). This indicates that in an acidic environment,

the retinal binding pocket is structurally disrupted within seconds after initial exposure to SDS but Schiff base hydrolysis is significantly slowed down. These observations are consistent with previous reports of bacteriorhodopsin in DMPC/CHAPSO micelles being unfolded at pH 1.2 in the presence of 6 mM SDS [276]. pSRII unfolded in SDS at pH 2.0 also showed less aggregation than at pH 6.0 (Figures 2.8, 2.10 & 2.11). Altogether, the results suggest that the attached retinal chromophore might be exerting a stabilising effect against protein aggregation. Hence, compared to at pH 6.0 (Figure 2.14), acid denaturation of pSRII at pH 2.0 (without SDS) led to very slow protein unfolding and Schiff base hydrolysis, and addition of SDS led to disruption of the retinal binding pocket within seconds but Schiff base hydrolysis is significantly slowed down.

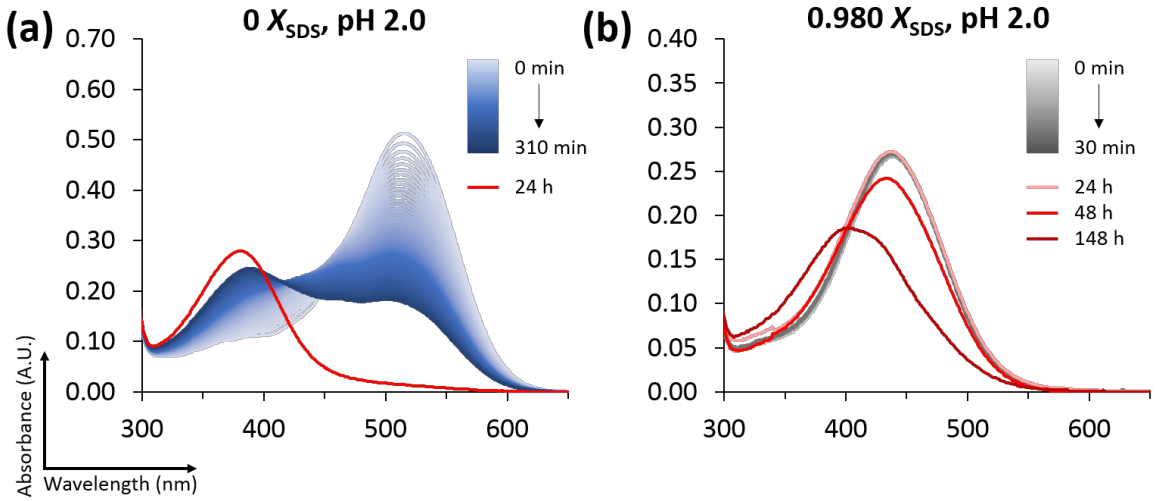


Figure 2.29: Time-resolved UV/vis spectra of pSRII unfolded in SDS at pH 2.0
(a) pSRII in 0 X_{SDS} at pH 2.0. Spectra were recorded at 5-min intervals. **(b)** pSRII in 0.980 X_{SDS} (3% SDS + 0.1% c7-DHPC) at pH 2.0. Spectra were recorded at 1-min intervals. In both panels, spectra recorded after prolonged periods of unfolding (≥ 20 h) are plotted in different shades of red.

2.3.6.3 Tertiary structure changes

Tryptophan fluorescence of pSRII in the absence of SDS at pH 2.0 also has λ_{max} of 337 nm, indicative of a similar environment surrounding the tryptophan residues as that at pH 6.0. The presence of 0.769 X_{SDS} at pH 2.0 led to the burial of tryptophan residues into a more hydrophobic environment, as evidenced by the spectral blue shift to 332 nm. Further addition of SDS led to a spectral red shift to 340 nm at 0.998 X_{SDS} (30% SDS in 0.1% c7-DHPC) as the tryptophan residues become more solvent-exposed (Figure 2.30a), similar to pSRII

unfolded in SDS at pH 6.0. Akin to the results at pH 6.0 (Figure 2.16), pSRII in SDS at pH 2.0 also showed higher fluorescence intensities with increasing X_{SDS}. Fluorescence intensities remained constant over 30 min (Figure 2.30b), consistent with the lack of tertiary structure changes over the same timescale observed by UV/vis spectroscopy (Figure 2.29b). On the other hand, fluorescence intensities increased slowly in 0 X_{SDS} at pH 2.0, consistent with the very slow protein unfolding and Schiff base hydrolysis in 0 X_{SDS} observed by UV/vis spectroscopy (Figure 2.29a).

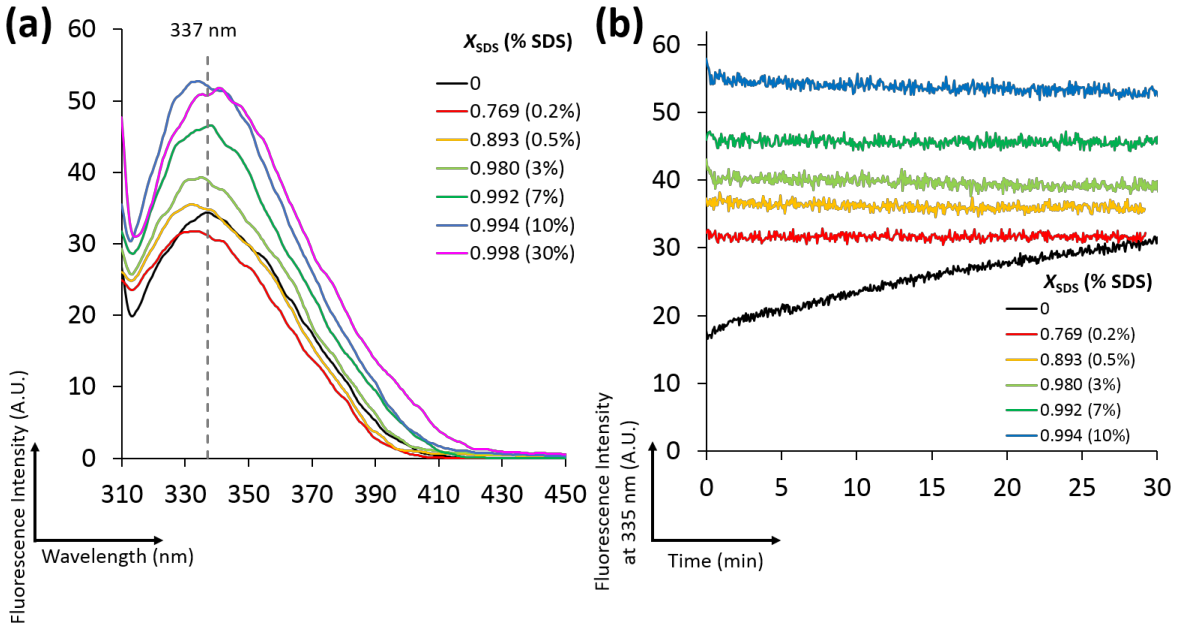


Figure 2.30: Tryptophan fluorescence of pSRII unfolded in SDS at pH 2.0
Emission spectra **(a)** and time-courses **(b)** of 7 µM pSRII unfolded for 30 min in 0–0.998 X_{SDS} (0–30% SDS) at pH 2.0. In panel a, the emission λ_{max} (337 nm) of pSRII in 0 X_{SDS} at pH 2.0 is shown using dotted line.

2.3.6.4 Solvent accessibility of the binding pocket

10 mM hydroxylamine at pH 2.0 led to near-complete formation of retinal oxime after 5 h (Figure 2.31a), indicating greater solvent accessibility of the retinal binding pocket than at pH 6.0. The loss of native chromophore in the presence of hydroxylamine at pH 2.0 was faster than in the absence of hydroxylamine (Figure 2.31c). Addition of SDS also led to fast denaturation such that only the SR_{440} intermediate was observed within the experimental dead-time (Figure 2.31b). However, the reaction did not reach completion even after 90 h (cf. complete reaction within 30 min at pH 6.0 shown in Figure 2.18b(ii)), as both retinal

oxime (360 nm) and retinal PSB (440 nm) were visible, manifesting as a very broad absorbance peak at ~410 nm. This suggests that the reaction of hydroxylamine with the retinal Schiff base is significantly slowed down at pH 2.0, likely due to a decrease in the reactivity of the hydroxylamine at low pHs [311]. The presence of hydroxylamine at pH 2.0 led to smaller loss of secondary structure within 30 min (Figure 2.27b), possibly due to incomplete Schiff base cleavage by hydroxylamine at pH 2.0 within this time frame, hence resulting in partial stabilisation of some elements of pSRII secondary structure as hypothesised above for pSRII molecules which still harbour PSB. The same unfolding transition point was observed, regardless of the absence or presence of hydroxylamine (Figure 2.27b). Fluorescence intensities increased very slowly over time, with slower rates being observed at higher X_{SDS} (Figure 2.32b), suggesting that SDS might create a 'shielding' effect at pH 2.0, either by occluding the access of hydroxylamine to the retinal binding pocket or by causing a decrease in the reactivity of hydroxylamine.

In summary, the reaction of hydroxylamine with retinal Schiff base at both pH 6.0 and pH 2.0 were dependent on the accessibility of hydroxylamine to the binding pocket, which increased in the presence of SDS. However, this reaction was significantly slowed down at pH 2.0 due to a decrease in the reactivity of hydroxylamine, leading to incomplete Schiff base cleavage and smaller secondary and tertiary structure changes.

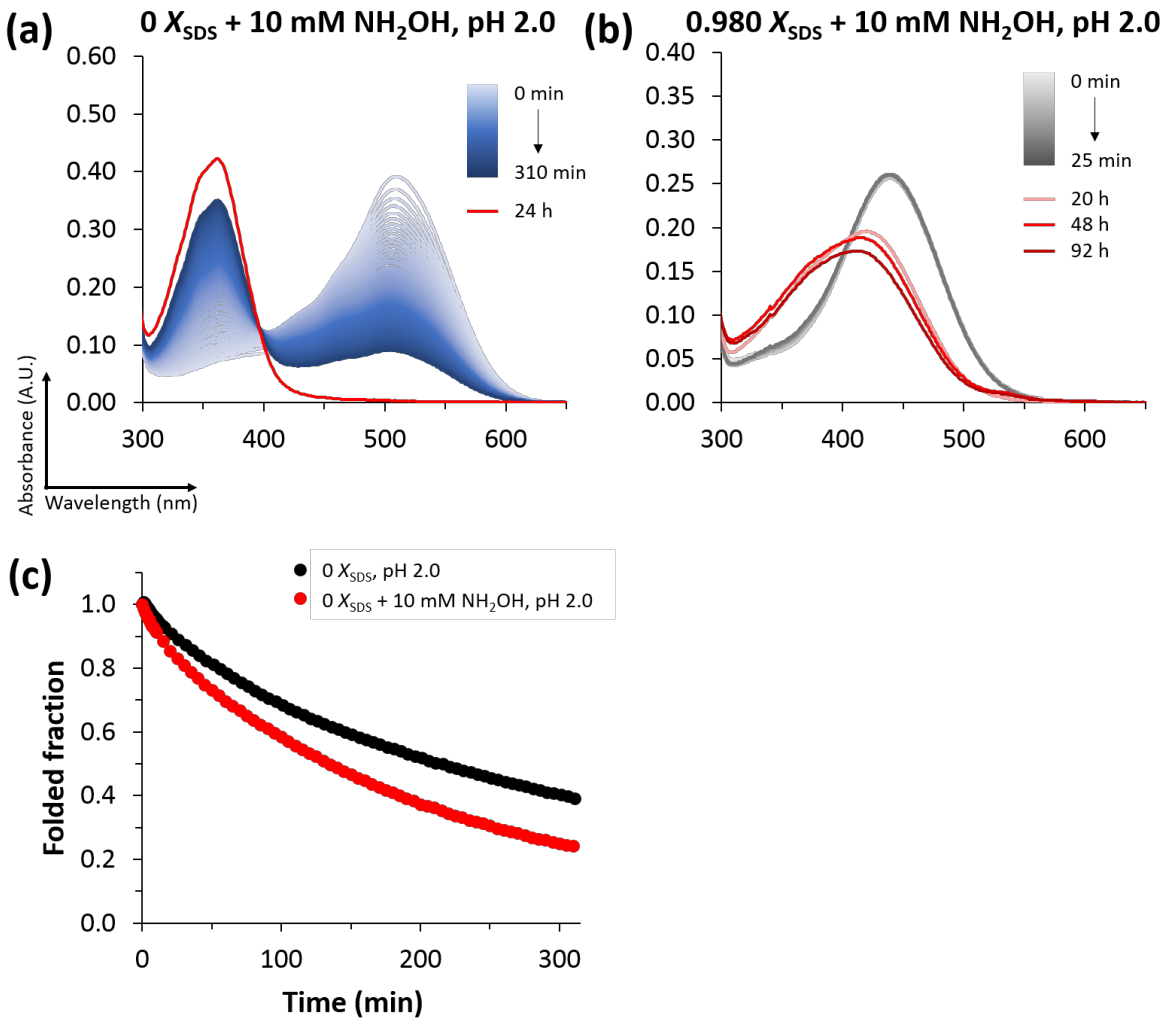


Figure 2.31: Time-resolved UV/vis spectra of pSRII unfolded in SDS at pH 2.0 with 10 mM hydroxylamine
(a) pSRII in 0 X_{SDS} + 10 mM hydroxylamine pH 2.0. Spectra were recorded at 5-min intervals. **(b)** pSRII in 0.980 X_{SDS} (3% SDS + 0.1% c7-DHPC) + 10 mM hydroxylamine, pH 2.0. Spectra were recorded at 1-min intervals. In both panels, spectra recorded after prolonged periods of unfolding ($\geq$ 20 h) are plotted in different shades of red. **(c)** Time-course for the decrease in folded fraction (monitored by the decay in absorbance at 498 nm) for pSRII in 0 X_{SDS} in the absence (black trace) or presence (red trace) of 10 mM hydroxylamine.

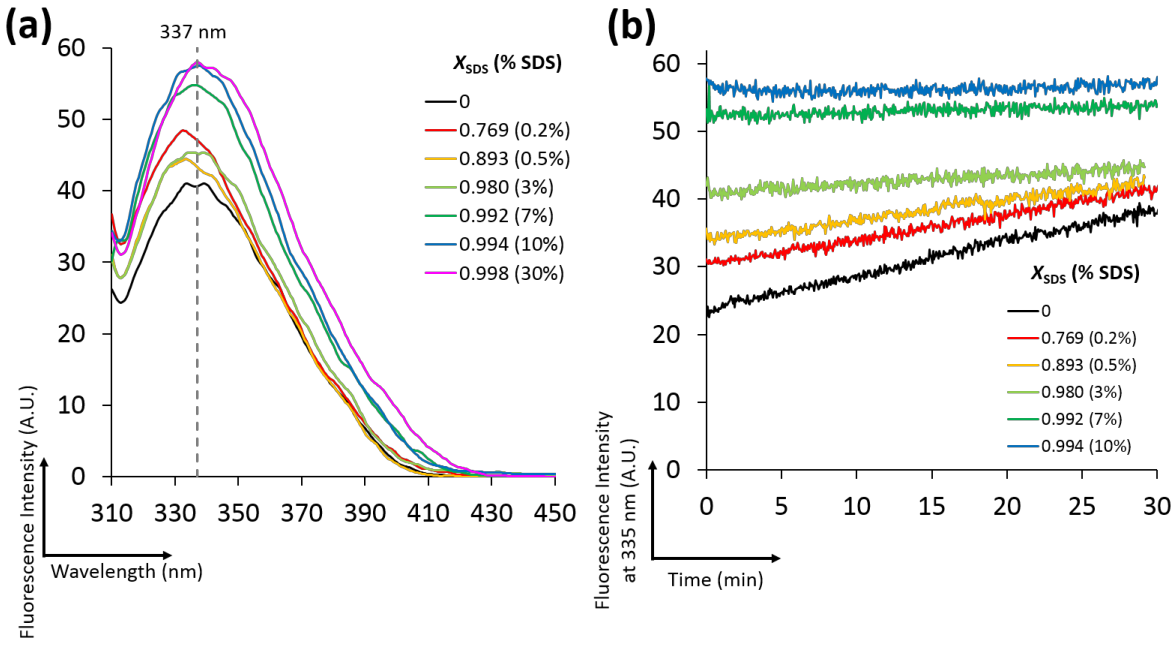


Figure 2.32: Tryptophan fluorescence of pSRII unfolded in SDS + 10 mM hydroxylamine at pH 2.0

Emission spectra **(a)** and time-courses **(b)** of 7 µM pSRII unfolded for 30 min in 0–0.998 X_{SDS} (0–30% SDS) + 10 mM hydroxylamine at pH 2.0. In panel a, the emission λ_{max} (337 nm) of pSRII in 0 X_{SDS} + 10 mM hydroxylamine at pH 2.0 is shown using dotted line.

2.4 Discussion

Membrane proteins are notoriously difficult to denature due to their inherent stability once functionally reconstituted into membrane mimetics and their tendency to aggregate if the hydrophobic mimetic environment were perturbed. The studies presented in this chapter first employed the screening of different denaturants and additives to simultaneously maximise the extent of unfolding in pSRII and minimise aggregation of the unfolding intermediates and unfolded states. SDS was identified as the most suitable condition for pSRII unfolding studies, causing non-specific aggregation only after prolonged exposure of the protein to SDS over several days. Using various biophysical techniques (CD, UV/vis spectroscopy, tryptophan fluorescence), the molecular properties of the denatured states were characterised within 2 h of SDS exposure, which is the minimum time required to fully denature pSRII across all X_{SDS} with minimal aggregation. SDS-denatured pSRII preserved much of its helical secondary structure while its tertiary structure was largely disrupted, as evidenced by loss of the retinal chromophore and changes in tryptophan fluorescence intensities. The kinetics of unfolding and refolding reactions were characterised, yielding the chevron plot and showing pSRII to be a model protein with a complex folding mechanism.

2.4.1 Denaturant screening

SDS was identified as the most suitable condition for unfolding pSRII. Non-specific aggregation was not observed within the typical two-hour timeframe for biophysical and kinetics studies, despite being visible in the form of soluble oligomers over much longer timescale of days. While some guidelines are available for minimising aggregation mostly in the contexts of globular proteins, intrinsically disordered proteins, and protein refolding [279,283,312], general recommendations for minimising aggregation of membrane proteins under denaturing conditions are lacking.

The motivation behind screening different denaturants and additives is to find conditions which favour unfolding and disfavour aggregation. Protein aggregation consists of several steps: (1) structural changes within monomers lead to the formation of partially-folded species with aggregation-prone regions being exposed to solvent; (2) nucleation and/or formation of oligomers; and (3) formation and growth of fibrils or amorphous aggregates

[313]. Structural changes in unfolded monomers are typically attributed to attraction between hydrophobic patches from different protomers to replace intramolecular contacts which have been disrupted during denaturation and to offer structural stabilisation. Multiple aggregation pathways might exist, yielding a complex ensemble of different partially-folded and aggregated structures. This chapter has explored the typical methods for shielding exposed hydrophobic surfaces and balancing surface charges, but the case of SDS-denatured pSRII proved to be more complex. This is a challenging problem because the same driving forces, namely formation of hydrogen bonds and burial of hydrophobic surfaces, are involved in both aggregation and protein folding [283]. A compromise was found in which aggregation was minimised without limiting the extent of unfolding by conducting biophysical and kinetics investigations at shorter time periods of 2 h.

Previous unfolding studies of bacteriorhodopsin and bovine rhodopsin also identify SDS as the most suitable denaturant, and give rise to many biophysical and structural studies on the SDS-denatured states of these two proteins. The striking similarities in the properties of bacteriorhodopsin and pSRII in the presence of different denaturants (urea, GdmCl, SDS mixed micelles at neutral and acidic pHs) suggest that the two proteins might have comparable unfolding mechanisms [276]. In contrast, the effects of denaturants on bovine rhodopsin are very different: urea and GdmCl both unfold rhodopsin but lead to aggregation, while SDS alone and 3% SDS + 8 M urea unfold rhodopsin to different extents, both without causing aggregation [270]. These differences in the responses to different denaturants and the presence of aggregates in pSRII but not rhodopsin upon denaturation in SDS suggest different folding/unfolding pathways of pSRII vs. rhodopsin.

2.4.2 SDS-denatured states of pSRII at pH 6.0

Far-UV CD, UV/vis absorbance spectroscopy, tryptophan fluorescence and kinetics studies were used in this chapter to characterise the SDS-denatured states of pSRII at different X_{SDS}. Structural changes in pSRII upon denaturation in SDS at pH 6.0 included an increase in accessibility of the retinal binding pocket with concomitant larger exposure of this region to solvent up to 0.943 X_{SDS}. Most of the transmembrane helical structure was preserved, and the small loss of α-helicity was attributed to fraying of helix ends. Beyond 0.994 X_{SDS}, non-native helical structure and burial of tryptophan residues were observed, likely because

of the helix-inducing effect of the cylindrical SDS micelle environment wrapping around the protein [119,308]. The unfolding transition was comparable across the three biophysical techniques. Extensive comparison of the unfolding rate constants extracted using different biophysical techniques was also done for pSRII unfolding in SDS + 0.1% c7-DHPC.

It is noticeable that pSRII unfolded much slower than both bacteriorhodopsin and bovine rhodopsin. This appears to contradict existing knowledge on the compactness of the native states of these proteins. Rhodopsin is the most compact and has limited susceptibility to bleaching even in the presence of 50 mM hydroxylamine [314], while micellar suspensions of monomeric bacteriorhodopsin [303] and pSRII (Figure 2.18b) showed detectable bleaching. While the utilisation of different detergent mixed micelle systems complicates a direct comparison of studies on the three proteins (c7-DHPC for pSRII, DMPC/CHAPS for bacteriorhodopsin, and DDM for bovine rhodopsin), this factor is unlikely to solely account for the large differences in unfolding rate constants. Preliminary studies of SDS-mediated unfolding of pSRII reconstituted in DDM yielded comparable rate constants for protein unfolding and Schiff base hydrolysis as those for pSRII in SDS/c7-DHPC (Figure 2.33).

Altogether, comparative studies on the SDS-denatured states of pSRII, bacteriorhodopsin and bovine rhodopsin point toward differences in stability, unfolding mechanisms and protein-SDS interactions among the three proteins. More systematic comparisons using comparable conditions and membrane mimetics will be necessary to make more assertive conclusions on differences in kinetic and thermodynamic parameters between these proteins. Further investigation into the properties of mixed micelles and high-resolution studies on protein-detergent interactions in SDS-denatured pSRII will provide further valuable insights into the molecular nature of the denatured states and the unfolding/folding pathway.

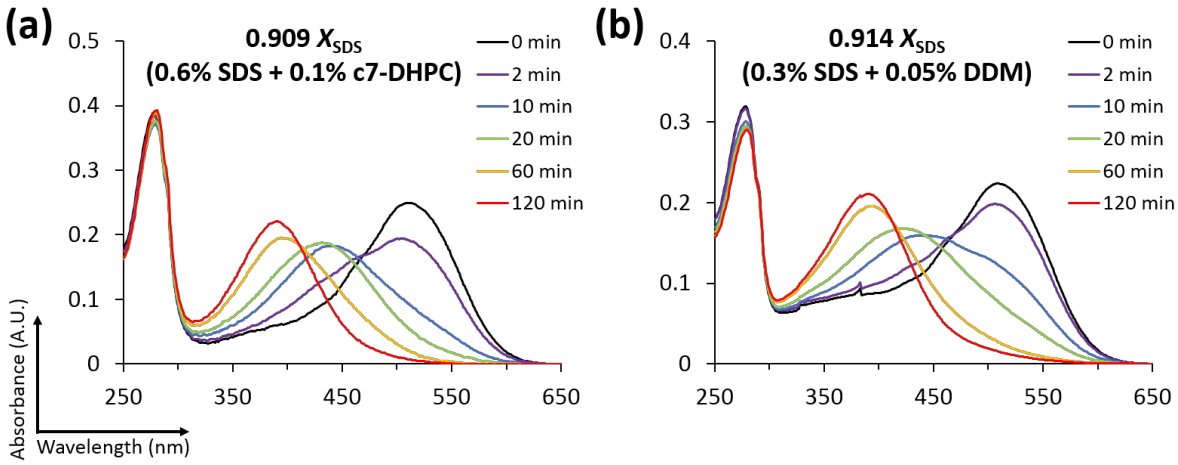


Figure 2.33: Time-resolved UV/vis spectra of pSRII unfolded in (a) 0.6% SDS + 0.1% c7-DHPC (0.909 X_{SDS}) and (b) 0.3% SDS + 0.05% DDM (0.914 X_{SDS})

2.4.3 SDS-denatured states of pSRII at pH 2.0

Acid denaturation of pSRII in 0 X_{SDS} led to small structural changes, as evidenced by the small increase in solvent accessibilities of retinal and tryptophan residues and also the slow bleaching of pSRII in the presence of hydroxylamine. Similarly, acid-denatured bacteriorhodopsin studied by oxidative labelling suggests the absence of large-scale unfolding [315], and the crystal structure reveals subtle changes in the retinal binding pocket and across the protein [316].

pSRII in SDS at pH 2.0 showed very fast loosening of the retinal binding pocket within seconds, but extremely slow retinal Schiff base hydrolysis on the timescale of days, rather than minutes to hours as in pH 6.0. The increase in solvent accessibility of tryptophan residues and bound retinal were also slowed down to timescales matching that of retinal Schiff base hydrolysis, whereas loss in α-helicity was accelerated to within minutes. In comparison to pH 6.0, a greater extent of α-helical content was preserved at pH 2.0. A detailed comparison between the apo-protein states (SO_{390}) in SDS at pH 2.0 and pH 6.0, however, would be complicated by changes in Schiff base reactivity at acidic pH and aggregation over prolonged time.

Although the presence of retinal in retinal-binding proteins enables convenient monitoring of unfolding and refolding by spectroscopic techniques, this also leads to complications with

further studies of SR_{440} due to fast protein conformational changes and very slow retinal Schiff base hydrolysis occurring simultaneously. SR_{440} in SDS at pH 2.0 was particularly interesting because its decay due to retinal Schiff base hydrolysis was significantly slowed down, enabling SR_{440} to remain observable for days rather than minutes. Further structural studies are required to understand to what extent SR_{440} at pH 6.0 and at pH 2.0 are structurally similar to each other, and whether SR_{440} at pH 2.0 could be used as a substitute for further structural characterisation.

It was also noted that pSRII in SDS at pH 2.0, which is predominantly SR_{440}, aggregated slower than pSRII in SDS at pH 6.0, which is predominantly SO_{390}. This suggests possible roles of the bound retinal in preventing the exposure of aggregation-prone surfaces and in the folding of pSRII. In fact, retinal binding is the rate-limiting step in bacteriorhodopsin folding [317]. More detailed structural studies will hopefully clarify whether there are protein conformational changes during retinal Schiff base formation in pSRII.

2.4.4 Folding model for pSRII

The main challenge for refolding kinetic studies is identifying solvent conditions in which refolding occurs with good efficiency (good yield of refolded material) and on a timescale which is amenable for biophysical techniques (seconds to minutes). Many studies of refolding kinetics in literature [18] and those presented in this chapter utilise a simple one-step dilution of the unfolded protein into a buffer devoid of denaturants. Additional detergents, lipids or additives may be introduced to facilitate refolding [18]. It is also important to establish assays for assessing the functionality of the refolded protein. For pSRII, reformation of the native chromophore monitored by UV/vis spectroscopy is indicative of a functional protein.

Although the SDS-denatured apo-protein states of bacteriorhodopsin and pSRII both retain significant amounts of secondary structure and no native retinal-protein interactions, there are fundamental differences in their unfolding/folding pathways and protein-SDS interactions. This is most notably exemplified by non-linearity of the folding arm in the pSRII chevron plot, suggesting either transient aggregation or the presence of “kinetically invisible” intermediate(s) in the folding of pSRII but not in bacteriorhodopsin. Given that there is no

dependence of the refolding rate constant on protein or detergent concentrations (Figure 2.23), transient aggregation and multiple micelle occupancy are unlikely, suggesting a complex folding pathway between SR_{440} and pSRII with on/off-pathway intermediate(s). It is difficult to reach a definitive conclusion about the existence and properties of the putative folding intermediate(s). Even for soluble proteins, interpretations of non-linear chevron plots and conclusions about the presence and properties of intermediate states are often elusive and require extensive ϕ-value analysis of different mutants, with the choice of mutants being carefully guided by the protein structure. Nevertheless, it is evident that the folding pathway between SR_{440} and folded pSRII, which predominantly involves the formation of native tertiary packing, shows clear deviations from the two-state model exemplified by bacteriorhodopsin.

Retinal binding was shown to be reversible, and the refolding yield from SO_{390} correlated with the amount of monomeric SO_{390} left after prolonged incubation in SDS over several days. Reversible unfolding of the apo-protein state can be achieved for bacteriorhodopsin and pSRII, both of which retain significant portions of α-helicity in SDS. While refolding of SO_{390} was not studied in considerable detail here, kinetic studies of refolding apo-bacterioopsin from SDS into lipid/detergent micelles have enabled detection of multiple folding intermediates in the folding pathway [120,132,317]. In comparison, despite extensive efforts, chemically-denatured rhodopsin, the apo-protein state of which is characterised by largely disrupted secondary and tertiary structures, has thus far resisted refolding [318]. Opsin formed upon bovine metarhodopsin II decay in DDM micelles is capable of rebinding 11-*cis* retinal to regenerate functional rhodopsin, but prolonged (up to 24 h) incubation in DDM leads to time-dependent increase in misfolding and decrease in regeneration yield, although the retinal-binding form of opsin can be reformed slowly when incubated in the presence of 11-*cis* retinal [319]. Refolding studies on these three proteins thus suggest that a minimal amount of correctly folded, stabilising structural elements might be required for reversible protein folding, and that the retinal chromophore might have roles in influencing the equilibrium between different apo-protein states or even the refolding of non-native apo-protein states. Once again, it must be emphasised that it is very challenging to obtain denatured states of membrane proteins which show minimal aggregation and are amenable to both biophysical characterisation and refolding studies.

2.5 Summary and further studies

By using various biophysical techniques, this chapter has led to the characterisation of the denatured states of pSRII in SDS/c7-DHPC mixed micelles and the kinetics of unfolding and refolding from both the unfolded state with a covalently-attached chromophore (SR_{440}) and the apo-protein state (SO_{390}). Similar to SDS-denatured bacteriorhodopsin, SDS-denatured pSRII retains a high α-helical content and loses most of its native retinal-protein contacts. The SDS-denatured states of pSRII, bacteriorhodopsin and bovine rhodopsin have different refolding properties, although refolding of bacteriorhodopsin and pSRII both involve predominantly the formation of native tertiary packing. Refolding of SDS-denatured bacteriorhodopsin with bound retinal adheres to two-state folding kinetics; the refolding kinetics of the equivalent SR_{440} state showed clear deviations from the two-state folding regime; and refolding of bovine rhodopsin remains elusive even from partially denatured states. It was also found that only monomeric SDS-denatured pSRII was capable of refolding from the apo-protein (SO_{390}) state to reconstitute the native chromophore.

This chapter presents a low resolution model for the unfolding and folding pathways of pSRII (Figure 2.34). Taking into consideration the different behaviours under unfolding and/or refolding conditions amongst the three retinal-binding proteins, pSRII is therefore proposed as a new model protein for membrane protein folding studies.

Further studies by techniques such as NMR (see Chapter 3) will be needed to (1) obtain residue-specific information on SDS-induced structural changes in pSRII and SDS-protein interactions; and (2) to probe for different species in the denatured states ensemble of pSRII. While CD is widely used for elucidating secondary structure changes, SDS has been reported to change the extinction coefficient of the peptide carbonyl group, leading to underestimation of the helical content of SDS-bound proteins (as is the case for bovine serum albumin) [320,321]. The α-helical content of SDS-denatured pSRII could be confirmed by measuring through-space nuclear Overhauser effects (NOEs), looking for sequential NH/NH NOEs. Although tryptophan fluorescence results clearly indicate increased solvent exposure of tryptophan residues upon SDS-mediated denaturation, it is difficult to determine the contribution from each of the six tryptophans since the fluorescence spectra are averages

across all tryptophans. Systematic studies on the dynamics and chemical environment of the backbone amide and side-chain indole of tryptophans by NMR could be achieved through [^{1}H,^{15}N]-correlated experiments or using fluorinated tryptophan analogues (see Chapter 3).

The conditions utilised in *in vitro* unfolding and refolding studies tend to be far removed from the conditions found in live cells. It is therefore important to explore methods which enable protein folding studies to be conducted in a biological context (see Chapter 5).

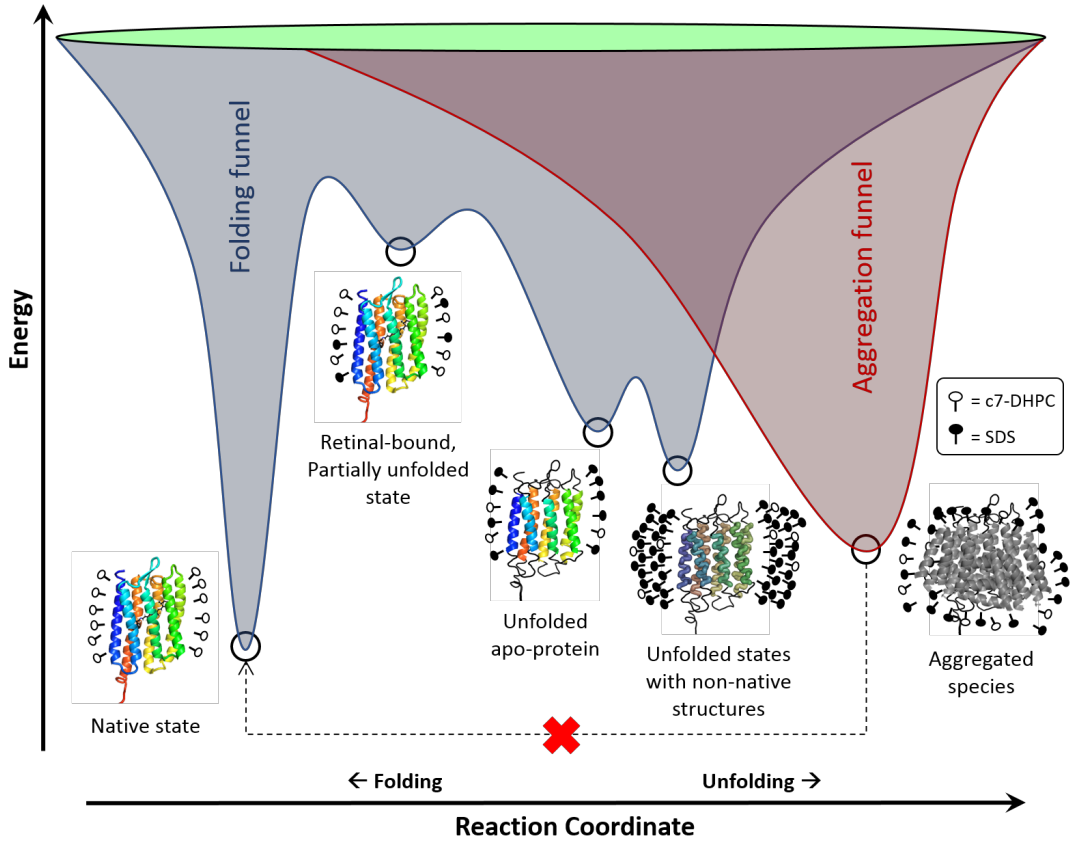


Figure 2.34: Summary model for the unfolding and folding pathways of pSRII
pSRII unfolds reversibly in SDS/c7-DHPC mixed micelles to form the retinal-bound, partially-unfolded SR_{440} state, followed by Schiff base hydrolysis to form apo-sensory opsin (SO_{390}). SO_{390} in high X_{SDS} has non-native structures due to the helix-inducing properties of SDS. Unfolded pSRII aggregates slowly over the timescale of days. The aggregated species cannot be refolded to yield the native chromophore.

3 Characterisation of denatured sensory rhodopsin II by solution-state NMR

3.1 Introduction

The structure and dynamics of denatured and unfolded proteins have been studied by various methods including single-molecule fluorescence spectroscopy, NMR, neutron scattering and computer simulations amongst other methods. Simulated thermal unfolding experiments suggest that pSRII is stabilised by a combination of mechanisms found for rhodopsin and bacteriorhodopsin [139], with a subset of helices breaking off early in the simulation and a core involving residues from other helices and loops remaining until late.

Solution-state NMR has established itself as one of the key methods for atomic resolution studies of the structure, function and dynamics of membrane proteins. Since unfolded proteins usually constitute an ensemble of different conformations, NMR is a highly suitable technique for characterising changes in structure and dynamics.

In this chapter, solution-state NMR was used for detailed characterisation of SDS-denatured pSRII. 1D ^{1}H and ^{19}F NMR were first used to characterise global changes in backbone amide protons and tryptophan side-chains (see Section 3.3.1). To understand residue-specific structural and dynamics changes, 2D [^{1}H,^{15}N]-SOFAST-TROSY-HSQC were recorded and changes in chemical shifts and peak intensities were analysed (see Sections 3.3.2–3.3.5). Having identified that changes in backbone amide dynamics is an important feature of denatured pSRII, ^{15}N relaxation experiments were performed to study backbone amide dynamics reflecting motions on different timescales, including fast fluctuations of NH bond vectors on the ps–ns timescale and exchange contribution on the ns timescale (see Section 3.3.6). In Sections 3.4–3.5, changes in chemical shifts and peak intensities were studied for acid-denatured pSRII and sensory opsin obtained by hydroxylamine-mediated bleaching of the retinal protonated Schiff base.

Overall, the NMR studies in this chapter have shed insight on differences in the unfolding pathways in the presence of different denaturants and the crucial role of the retinal chromophore in governing the structural integrity and dynamics of the pSRII helical bundle.

3.2 Theory: Nuclear Magnetic Resonance (NMR)

3.2.1 Overview

For all forms of spectroscopy, the spectrum arises from transitions between different energy states of a molecule. NMR spectroscopy arises from the fact that nuclei have a property known as spin, an intrinsic form of angular momentum carried by elementary particles and atomic nuclei. In the presence of an external magnetic field, the degeneracy of the nuclear spin energy levels is broken, allowing transitions between these energy levels. The ground and excited states in NMR arise from the interaction between the external magnetic field and a nuclear magnetic dipole moment, which describes the magnetic strength and orientation of a magnetic field generated by the nucleus in the presence of an external magnetic field. Nuclei with an odd mass number (e.g. ^{1}H, ^{13}C, ^{15}N, ^{31}P) have spin angular momentum ($\vec{S}$) due to the presence of an unpaired proton. All nuclei with an even mass number and an odd charge (e.g. ^{2}H, ^{14}N) also have spin angular momentum.

In the absence of an applied magnetic field, a collection of nuclear spins will adopt all possible orientations of the magnetic dipole. Once the spins are placed in a magnetic field, the nuclear magnetic moment and the applied field interact. The direction of the applied field is defined as the *z*-axis. The energy of the interaction depends on the angle between the magnetic moment and the applied field. The lowest energy arrangement is when the magnetic moment is parallel to the field, and the highest energy is when the magnetic moment is opposed to the field. There is an energetic preference for more spins to be aligned with the applied magnetic field, leading to the build-up of a bulk magnetisation vector which is in equilibrium with the applied field and pointing along the *z*-axis. The process by which the spins come to equilibrium in a magnetic field is called relaxation (see Section 3.2.5).

To obtain transitions between different nuclear spin states, radiofrequency (RF) waves with energy matching the difference in energy levels of these spin states must be applied. Hamiltonian operators are operators which represent the observable quantised energy in quantum mechanics. For a nuclear spin in a magnetic field of strength B_0 applied along the *z*-axis, the energy of interaction between the spin and the magnetic field is:

$$\hat{H} = -\gamma B_0 \hat{I}_z \tag{3.1}$$

where γ is the gyromagnetic ratio and $\hat{I}_z$ is an operator which represents the *z*-component of the nuclear spin angular momentum.

The eigenvalues of the Hamiltonian are the energy levels available to the system [199]. $\hat{I}_z$ has $(2I + 1)$ eigenfunctions, where *I* is the spin quantum number which is specific to a given nucleus, meaning the nuclear spin angular momentum is quantised. Each eigenfunction (ψ_m) of the operator $\hat{I}_z$ is characterised by m, another spin quantum number, which can only take values from $-I$ to *I* in integer steps. For spin-1/2 nuclei $\left(I = \frac{1}{2}\right)$, $\hat{I}_z$ has two eigenfunctions characterised by $m = +\frac{1}{2}$ or $-\frac{1}{2}$:

$$\hat{I}_z \psi_{\pm 1/2} = m\hbar\psi_{\pm 1/2} \tag{3.2}$$

The eigenfunctions of $\hat{I}_z$ are also eigenfunctions of the Hamiltonian in Equation 3.1, with eigenvalues:

$$E_m = -m\hbar\gamma B_0 \tag{3.3}$$

The two energy levels of a spin-half nuclei in a magnetic field are:

$$E_\alpha = -\frac{1}{2}\hbar\gamma B_0 \quad \text{and} \quad E_\beta = +\frac{1}{2}\hbar\gamma B_0 \tag{3.4}$$

where E_α is the ground state and E_β is the excited state by convention.

The energy of the allowed transition from the ground (α) state to the excited (β) state is:

$$\Delta E_{\alpha\to\beta} = E_\beta - E_\alpha = \hbar\gamma B_0 \tag{3.5}$$

The frequency corresponding to this energy gap is therefore:

$$v_{\alpha\to\beta} = \Delta E_{\alpha\to\beta}/h = \gamma B_0/2\pi \tag{3.6}$$

The Larmor frequency (v_0) is defined as:

$$v_{\alpha\to\beta} = -v_0 \tag{3.7}$$

For a single spin-1/2 nucleus, there is one allowed transition which results in a line in the spectrum at $-v_0$.

To record a spectrum, the magnetisation vector must be moved away from the *z*-axis and towards the *xy* (transverse) plane. To achieve this, a very small radiofrequency field (B_1) is applied along the transverse plane. Since the magnetisation is no longer aligned with the B_0 field, it would precess about the B_0 field at the Larmor frequency, sweeping out a cone with a constant angle β. Even though B_0 is many times stronger than B_1, the magnetisation would move away from the *z*-axis provided that an on-resonance pulse is applied (i.e. the transmitter frequency is the same as the Larmor frequency).

This excitation field can be applied to the sample in two ways, either by scanning through wavelengths (continuous wave NMR) or as a short burst of high power RF to excite a broad range of transitions (pulsed NMR) [322]. The most simple pulsed NMR experiment is the pulse-acquire experiment (Figure 3.1a). Equilibrium magnetisation first builds up along the *z*-axis. The short 90°(x) RF pulse rotates this magnetisation onto the –*y*-axis. The magnetisation precesses at the Larmor frequency and cuts the receiver coil, which is wound around the sample with its axis aligned in the *xy*-plane, thereby inducing a current in the coil. This current is then amplified, sampled at discrete times and digitised to give the free induction decay (FID) (Figure 3.1b).

This linearly oscillating FID can be decomposed into two counter-rotating fields. By adopting a rotating frame which rotates at the same frequency and in the same sense as the Larmor precession, the apparent Larmor frequency becomes zero. The precessing magnetisation is detected in the rotating frame. The observed frequency, Ω, is $\omega_s - \omega$, where ω_s is the precessional frequency of the spin and ω is the rate of rotation of the coordinate frame (or equivalently, the frequency of the applied B_1 pulse).

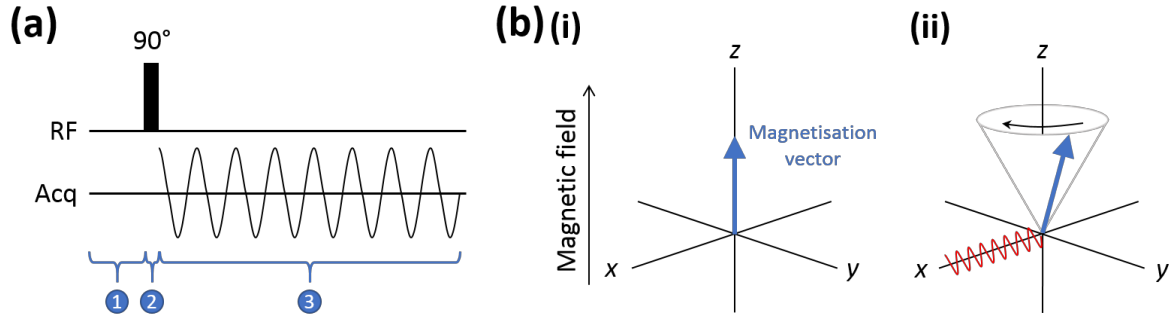


Figure 3.1: Pulse sequence (a) and schematic diagram showing changes in the magnetisation vector (b) for the pulse-acquire experiment.
(a) The experiment comes in three periods: 1 – the sample is allowed to reach equilibrium; 2 – a 90° pulse is applied to rotate the magnetisation into the *xy* plane; 3 – the RF power is switched off and signal is detected as it arises from the magnetisation while rotating in the *xy* plane. **(b) (i)** At equilibrium, a sample has net magnetisation along the direction of the magnetic field (*z*-axis) represented by the magnetisation vector (blue arrow). **(ii)** When the magnetisation vector is tilted away from the *z* axis, it executes a precessional motion, sweeping out a cone of constant angle and cutting a coil (red wave) wound around the *x* axis, thereby inducing a current in the coil. Figures adapted from [199].

Effectively, in the rotating frame, the magnetisation precesses in the transverse plane at the offset Ω, rotating through an angle Ωt during time *t*. The *x*- and *y*-components of the magnetisation evolve as follows:

$$M_x = M_0 \sin(\Omega t)\, e^{-t/T_2} \quad \text{and} \quad M_y = -M_0 \cos(\Omega t)\, e^{-t/T_2} \tag{3.8}$$

where Ω is the resonance frequency in the rotating frame and e^{-t/T_2} represents the decay of the excited state due to relaxation, with a time constant of T_2 (see Section 3.2.5).

Fourier transformation of the FID converts time-domain data into frequency-domain data, yielding a single line in the spectrum at minus the Larmor frequency. Peaks in NMR spectra typically have Lorentzian lineshape.

3.2.2 1D NMR

A plot of intensity against resonance frequency is known as a 1D NMR spectrum. Resonance frequencies are typically reported as 'chemical shifts' in units of ppm (parts per million), which accounts for the measured frequencies (in MHz) relative to a reference frequency and the strength of the applied magnetic field.

The resonance frequency of a nuclear spin depends on the magnetic field strength at the nucleus. This magnetic field usually differs slightly from the applied field because the magnetic field at each nucleus is shielded by the electron density surrounding it. The electron density depends on the chemical environment of the nuclear spin, for example the presence or absence of electronegative atoms in close proximity. As the electron density around the nucleus increases, the effective field decreases, leading to lower resonance frequencies (chemical shifts). Chemical shifts differ between atoms and between molecules.

In proteins, amide protons resonate at ~ 8 ppm, H_α protons at ~ 4 ppm and methyl protons (e.g. H_β and others) at ~ 1 ppm (Figure 3.2a). The amide proton has a high chemical shift because the nitrogen atom is more electron-withdrawing than carbon. The reduced electron density at the amide proton decreases the shielding and therefore increases the effective field and resonance frequency. Similarly, the H_α shifts are higher than the methyl H shifts and tryptophan side-chain H_ε chemical shifts are ~ 10 ppm because of their proximity to the electronegative nitrogen. Similar ordering of chemical shifts can also be observed in carbon atoms in proteins. Nitrogen atoms also have distinct chemical shifts (Figure 3.2b). For example, glycine ^{H}N resonate at ~ 110 ppm whereas tryptophan ^{H}N resonate at ~ 120 ppm, and tryptophan side-chain N_ε at ~ 130 ppm.

In practical NMR spectroscopy, several resonances are usually expected in the spectrum, each of which has a different Larmor frequency. A separate magnetisation vector can be associated with each detected NMR-active nucleus, and the detected signal will be the sum of contributions from each vector. Given the modest range of chemical shifts, even a small protein of around 50 residues will yield a complex spectrum containing many overlapping peaks. One solution to this problem is to increase the dimensionality of the NMR experiment, so that peak positions are defined by two or more resonance frequencies. Multi-dimensional NMR along with other techniques (see Section 3.2.3) have been applied to resolve spectra for proteins in the 30–50 kDa range.

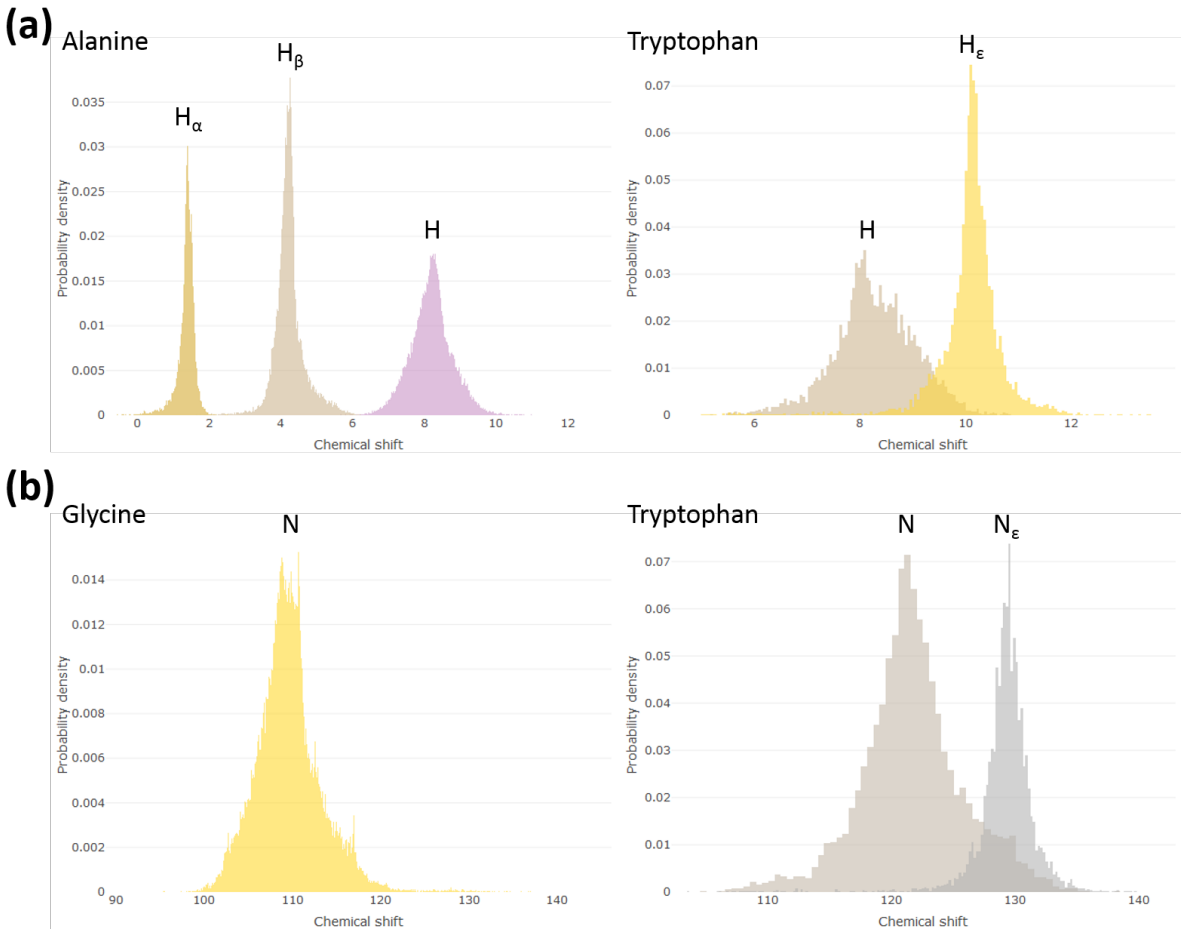


Figure 3.2: ^{1}H and ^{15}N chemical shift distribution from the BMRB
Distribution of ^{1}H chemical shifts of Ala and Trp residues **(a)** and ^{15}N chemical shifts of Gly and Trp residues **(b)** reported on the Biological Magnetic Resonance Bank (BMRB). Note that distributions of Trp side-chain H_ε and N_ε are also shown. Plots are generated using the RBMRB package on R.

The presence of multiple signals in the spectrum indicates that not all nuclei can be resonating at the same frequency as the B_1 field, meaning these signals are off-resonance. With a modern spectrometer, the B_1 field strength is sufficiently large such that the effective field for a range of resonances lies very close to the x-axis, although minor off-resonance excitation outside the bandwidth and phase errors could still affect measurements which require very high sensitivity.

For ^{1}H NMR on aqueous samples, solvent suppression is typically required as the water signal can be several orders of magnitude stronger compared to the signals of interest due to

the presence of non-deuterated protons. The solvent signal needs to be selectively perturbed such that the associated magnetisation is not in the transverse plane during signal detection.

3.2.3 2D NMR

The motivation of multidimensional NMR are to improve signal separation, and to filter out certain signals through their connections to other nuclei. One of the key features which differentiates NMR from many other forms of spectroscopy is that the excited states are relatively long lived, with lifetimes in the ms–s range. Consequently, the excited state can be manipulated to pass magnetisation from one nucleus to another. The frequencies can be measured for each of the nuclei through which magnetisation is passed. As a result, signals which correlate the frequencies between multiple nuclei can be obtained. In such correlated spectra, each transfer can be visualised as an independent nuclear frequency dimension (axis). This forms the basis of multi-dimensional NMR. Signals in the spectra occurring at the intersection of two or more frequencies indicate a correlation between the corresponding nuclei.

The HSQC (heteronuclear single-quantum coherence) experiment is widely used for recording one-bond correlation spectra for covalently bonded ^{1}H-^{15}N groups and ^{1}H-^{13}C groups. These functional groups are abundant in molecules of biological interest (peptides, proteins and nucleic acids). Given the low natural abundance of ^{13}C and ^{15}N isotopes, isotopically-labelled recombinant proteins can be produced by using minimal medium containing nutrients enriched with the desired isotope, for example using $^{15}NH_4Cl$ as the nitrogen source and ^{13}C-glucose as the carbon source for *E. coli* cultures.

In the 2D [^{1}H,^{15}N] HSQC experiment (Figure 3.3), magnetisation is transferred from ^{1}H spin to the covalently bonded ^{15}N spin. The ^{15}N spin magnetisation evolves for t_1, during which it acquires a frequency (offset) label (the chemical shift evolves). Finally, this magnetisation is transferred back to ^{1}H, where it is observed. The resulting spectrum has peaks centred at the offset of the ^{1}H spin in the ω_1 dimension and at the offset of the ^{15}N spin in the ω_2 dimension.

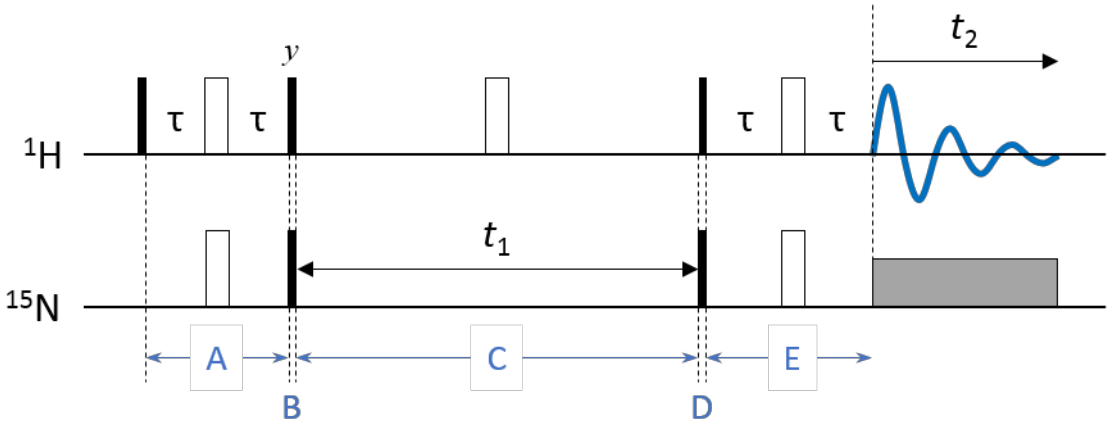


Figure 3.3: Basic pulse sequence for the 2D [^{1}H,^{15}N] heteronuclear single quantum correlation (HSQC) experiment
The pulse sequence starts with equilibrium magnetisation on the ^{1}H spin, which is transferred to the ^{15}N spin using an INEPT sequence (periods A–B). The ^{15}N magnetisation evolves during the t_1 period, with the centrally placed 180° ^{1}H pulse refocusing the evolution of the one-bond $^1J_{NH}$ coupling (period C). This magnetisation is transferred back to the ^{1}H spin (periods D–E), where it is observed in the presence of broadband decoupling (grey box). 90° pulses are represented by black lines and 180° pulses are represented as white boxes. Pulses are applied with phase *x* unless indicated otherwise.

The 2D [^{1}H,^{15}N] HSQC spectrum shows a signal for each covalently bonded ^{1}H-^{15}N group. The spectrum is plotted like a topographic contour map. Each signal has an intensity, a ^{1}H chemical shift and a ^{15}N chemical shift. For proteins, the ^{15}N-HSQC spectrum contains one peak for each backbone amide proton except those of Pro residues, a peak for each indole side-chain of Trp residues, and pairs of peaks for the side-chain amide groups of Asn and Gln residues. Signals from the $H_\varepsilon N_\varepsilon$ guanidine groups of arginine may also be observed. In essence, the ^{15}N-HSQC provides a high-resolution fingerprint of the amide backbone and side-chain atoms of selected residues of a protein.

Several factors affect the intensities and dispersion of signals in the [^{1}H,^{15}N] HSQC, including whether the protein is folded, the size of the protein and its dynamics.

Completely unfolded proteins show poor signal dispersion since all amide protons are exposed to the solvent and are in similar chemical environments. Residues in proteins which are rich in β-sheets are exposed to a wider variety of chemical environments, and therefore give rise to better signal dispersion than proteins which are rich in α-helices.

Narrow linewidths require long-lived excited states in NMR experiments. Factors which lead to rapid decay of excited states back to the ground state via relaxation give rise to broader lines (see Section 3.2.5). Folded proteins give well-defined signals. Completely unfolded proteins tend to exhibit sharp lines. Partially folded proteins often give rise to poor spectra quality. Interconversion of a protein between different conformations on the μs–ms timescale can cause additional line broadening due to chemical exchange (see Section 3.2.4).

Slow molecular reorientation is also a major cause of rapid signal decay and broad lines. This is usually apparent for samples with large molecular size, or when the molecule of interest interacts with binding partners, oligomerises or aggregates.

3.2.4 Titration and chemical exchange

Solution-state NMR spectroscopy is a powerful tool for analysis of structural and dynamical aspects of biomolecular interactions and equilibria. Systematic studies of changes in chemical shifts and lineshapes of NMR signals upon titration of an interaction partner can yield valuable structural, thermodynamic and kinetic information on the interaction. The appearance of resonances in NMR spectra during a titration experiment depends on the number of different species (e.g. free and bound forms) and the rate of exchange between these species relative to the frequency difference between them.

The exchange between two states, A and B, can be described as:

$$A \underset{k_B}{\overset{k_A}{\rightleftharpoons}} B \tag{3.9}$$

There are three distinguished exchange regimes for a two-state exchange system (Figure 3.4). In the slow exchange regime, the exchange rate (k) is much smaller than the frequency difference ($\Delta\omega$) between the two species ($k << \Delta\omega$). In the fast exchange regime, where $k >> \Delta\omega$. Slow exchange gives rise to two resonances centred at ω_A and ω_B, the chemical shifts of species A and B, respectively. The intensity of each peak reflects the concentration of the corresponding species. Fast exchange allows the observation of a single resonance with an average chemical shift δ_{obs} positioned at the population weighted average of the two resonance frequencies:

$$\delta_{obs} = p_A\delta_A + p_B\delta_B \tag{3.10}$$

where p and δ are the populations and chemical shifts of the two species.

When the proportions of the two species are varied, for example in a titration experiment, the signal will move gradually from the exchange-free chemical shift of one species to that of the other.

In the intermediate exchange regime, where k and Δω are comparable, signals broaden considerably and shift at the same time in a titration.

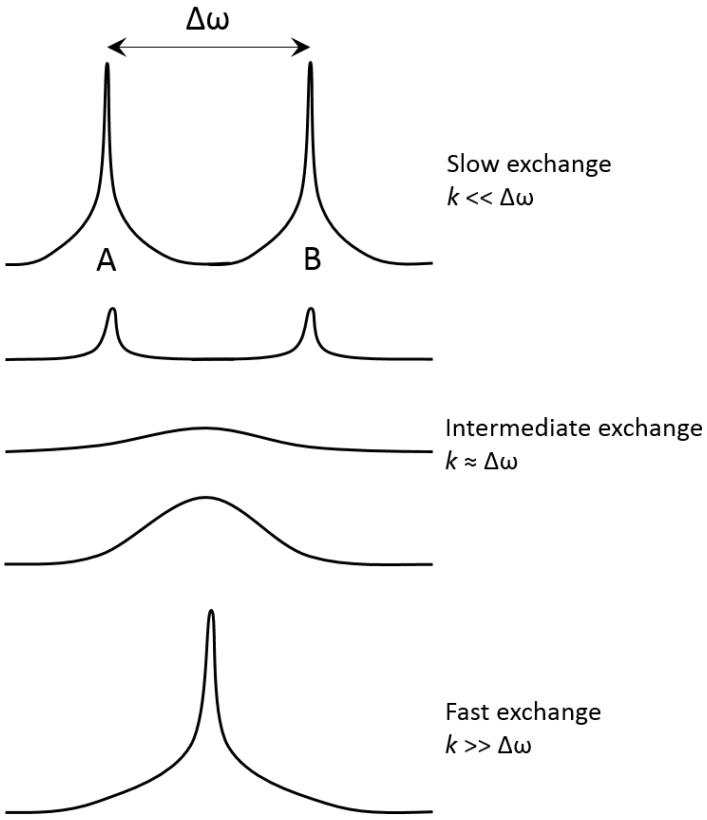


Figure 3.4: Chemical exchange regimes
Schematic diagrams showing the effects of chemical exchange between two species, A and B, assuming the populations of A and B are equal. In the slow exchange regime, where the rate of exchange is much smaller than the frequency separation between the two species (k << Δω), two lines are observed. As the exchange process accelerates so that k becomes similar to the frequency separation (k ≈ Δω), the lines first broaden and merge into one (intermediate exchange regime). Further increase in the exchange rate constant (k >> Δω) causes the merged line to narrow. Figure adapted from [199].

The populations of species A and B and the corresponding exchange rates, k_A and k_B, are often different, leading to asymmetric exchange. The consequences are: in the slow exchange regime, peak broadening of the minor state becomes much more significant; in the fast exchange regime, the population-weighted average peak position would be more biased towards the chemical shift of the major state. These factors further complicate spectral analysis and identification of chemical exchange regimes.

In a HSQC titration experiment with exchange between two species, an NH signal in superimposed HSQC spectra would move in a straight line. Nonlinear plots indicate the presence of multiple species and the titration affects the chemical shifts of each species differently. The presence of multiple species could be due to multiple binding interactions or multiple conformational changes [323].

$$P + L \leftrightarrow P^* + L \leftrightarrow PL \tag{3.11}$$

$$P + L \leftrightarrow PL \leftrightarrow P^*L \tag{3.12}$$

where P is the protein, P* is the protein with conformational change(s) and L is an interacting partner (e.g. non-specific or specific small molecule, ligand, binding partner).

In the context of protein unfolding, interactions of the denaturant with folded protein, folding intermediates and unfolded protein could be very different, leading to possible observation of different exchange rates (broadened signals) and nonlinear changes in chemical shifts. Hence, spectral analysis becomes increasingly challenging with multiple states in exchange and increasing numbers of exchange mechanisms, which often give rise to nonlinear titrations with broadened signals, leading to intractable changes in chemical shifts and signal intensities.

3.2.5 Molecular motion and nuclear spin relaxation

Relaxation is the process by which, over time, the bulk magnetisation returns to its equilibrium position. At equilibrium, there is z-magnetisation along the direction of the applied magnetic field and no transverse magnetisation. Transverse relaxation drives the transverse magnetisation to zero and longitudinal relaxation drives the longitudinal magnetisation to a particular steady state value.

Three parameters are typically measured for each nucleus to assess the dynamics of a protein: the longitudinal (R_1) and transverse (R_2) relaxation rate constants and the magnitude of the NOE between each amide proton and its attached nitrogen. R_1 and the $\{^1H\}$-^{15}N NOE report directly on the existence of motions on the ps–ns timescale, whereas R_2 additionally depends on slower motions on the µs–ms timescale. These data can be analysed to separate contributions from internal motion and overall Brownian motion. Internal motion on the ps–ns timescale is described by the generalised order parameter, S^2, often called an 'entropy meter', which describes the rigidity of each residue and it can have a value between zero (nucleus undergoing completely unrestricted motion) and one (nucleus moves only with the whole molecule).

3.2.5.1 Relaxation mechanisms

Depending on sample conditions, there are several molecular mechanisms which drive relaxation, including chemical shift anisotropy (CSA) and dipole-dipole (DD) interactions. Relaxation by DD interactions is due to coupling between the observed nucleus and the magnetic dipole of another spin. For example, a local magnetic field at the ^{15}N nucleus is produced by the magnetic moment of the attached ^1H nuclear spin. As the molecule rotates, the effect from the field of the attached ^1H nucleus varies in an orientation-dependent manner. The DD interaction also depends on the gyromagnetic ratio (γ) and the distance (r) between the two attached spins:

$$D^{NH}(t) \propto \gamma_N \gamma_H \, (3\cos^2\theta - 1)/r_{NH}^3 \tag{3.13}$$

With the r^3 distance dependence, the DD effect is only effective over very short distances (< 5 Å between two protons).

Chemical shift anisotropy arises from the fact that chemical shifts are determined by local magnetic fields, and asymmetric electron distributions lead to orientation-dependent changes in local fields experienced by the nuclei:

$$c = \gamma B_0(\sigma_\parallel - \sigma_\perp) \tag{3.14}$$

where $(\sigma_\parallel - \sigma_\perp)$ represents the difference between the parallel and orthogonal components of the chemical shift anisotropy tensor and B_0 is the external magnetic field. Tumbling in liquid samples usually averages this effect, producing an isotropic chemical shift in the spectra. The local field depends on the orientation of the molecule (anisotropic), meaning it is not

necessarily parallel to the applied field. Molecular reorientation modulates the local magnetic field and can act as a source of relaxation, provided that the frequencies of the modulating fields are in the range of those required for NMR transitions to occur. The extent to which the local field varies as the molecule tumbles depends on the anisotropy of the chemical shift (the extent to which the shift varies with orientation).

The ^{15}N nucleus in an N–H group is relaxed by two mechanisms: DD interaction with the attached ^{1}H and CSA of the ^{15}N itself. In both cases, the local field experienced by the ^{15}N fluctuates with rotational reorientation of the molecule due to thermal motion. Such motion alters the orientation of both the N–H vector and the ^{15}N shift anisotropy tensor with respect to the applied field. Variations of local fields arising from these two mechanisms must be correlated as they are modulated by the same motion. This is known as cross correlation between CSA and DD relaxation mechanisms. Cross correlation results in different relaxation rate constants and linewidths for different lines of the ^{15}N doublet. This effect is particularly pronounced for ^{15}N–^{1}H pairs in large molecules when the spectra are recorded at high field. For the studies in this chapter, TROSY experiments enable detection of only the ^{15}N–^{1}H multiplet component for which DD and CSA have almost cancelled, yielding narrow linewidths. Further detailed description of cross correlation and its effects on coupling constant measurements are given in Chapter 4.

3.2.5.2 Correlation function and spectral density function

In order to be effective at causing longitudinal relaxation, the local field must be oscillating at/close to the Larmor frequency. The correlation function is a way of characterising the time dependence of the random motion in a sample, and hence determining how much of the motion is present at the Larmor frequency. For a particular spin *i* experiencing a local field $B_{loc,i}(t)$ which varies in time, at a later time, the local field is expressed as $B_{loc,i}(t+\tau)$. The local field varies due to thermal motion in the sample, so $B_{loc,i}(t)$ is a random function of time. The correlation function $G(t,\tau)$ is defined as the average of the product $B_{loc,i}(t)B_{loc,i}(t+\tau)$ across all spins in the sample:

$$G(t,\tau) = \frac{1}{N}\sum_{i=1}^{N} B_{loc,i}(t)B_{loc,i}(t+\tau) = \overline{B_{loc}(t)B_{loc}(t+\tau)} \qquad (3.15)$$

where *N* is the number of spins in the sample.

Since the correlation function is a stationary random function (i.e. does not depend on the point *t* from which time is measured but only depends on the time interval τ, the function can be simplified as $G(\tau)$. $G(\tau)$ starts at a maximum at τ = 0, and decays exponentially with decay rate determined by the correlation time τ_c:

$$G(0) = \overline{B_{loc}(t)B_{loc}(t)} = \overline{B_{loc}^2} \tag{3.16}$$

$$G(\tau) = \overline{B_{loc}^2}\exp(-|\tau|/\tau_c) \tag{3.17}$$

The correlation is a function of time. Fourier transformation of the correlation function yields the spectral density function, *J*(*ω*), which is dependent on frequency and from which the amount of motion that is at the Larmor frequency can be determined.

$$J(\omega) = \overline{B_{loc}^2}\frac{2\tau_c}{1+\omega^2\tau_c^2} \tag{3.18}$$

$J(\omega)$ has its maximum value of $2\overline{B_{loc}^2}\tau_c$ at $J(0)$ (i.e. when ω = 0), and decays steadily as ω increases, with the decay rate being determined by τ_c (Figure 3.5a). An important feature of the spectral density is that the area under the curve is independent of τ_c.

$$J(\omega) = \int_0^\infty \overline{B_{loc}^2}\frac{2\tau_c}{1+\omega^2\tau_c^2}d\omega = \pi\overline{B_{loc}^2} \tag{3.19}$$

Hence, as τ_c becomes smaller, $J(0)$ decreases and at the same time the spectral density spreads out to higher frequency (see Figure 3.5a).

As previously mentioned, longitudinal relaxation is driven by modulation of the local field at the Larmor frequency, and the rate of longitudinal relaxation is therefore determined by $J(\omega_0)$, the spectral density at the Larmor frequency.

$$J(\omega_0) = \overline{B_{loc}^2}\frac{2\tau_c}{1+\omega_0^2\tau_c^2} \tag{3.20}$$

A plot of $J(\omega_0)$ against τ_c (Figure 3.5b) shows that $J(\omega_0)$ is at its maximum when $\tau_c \sim 1/\omega_0$. This indicates that the rate of longitudinal relaxation will also be at its maximum when $\tau_c \sim 1/\omega_0$. Correlation times which are shorter or longer than $\tau_c = 1/\omega_0$ will yield slower relaxation.

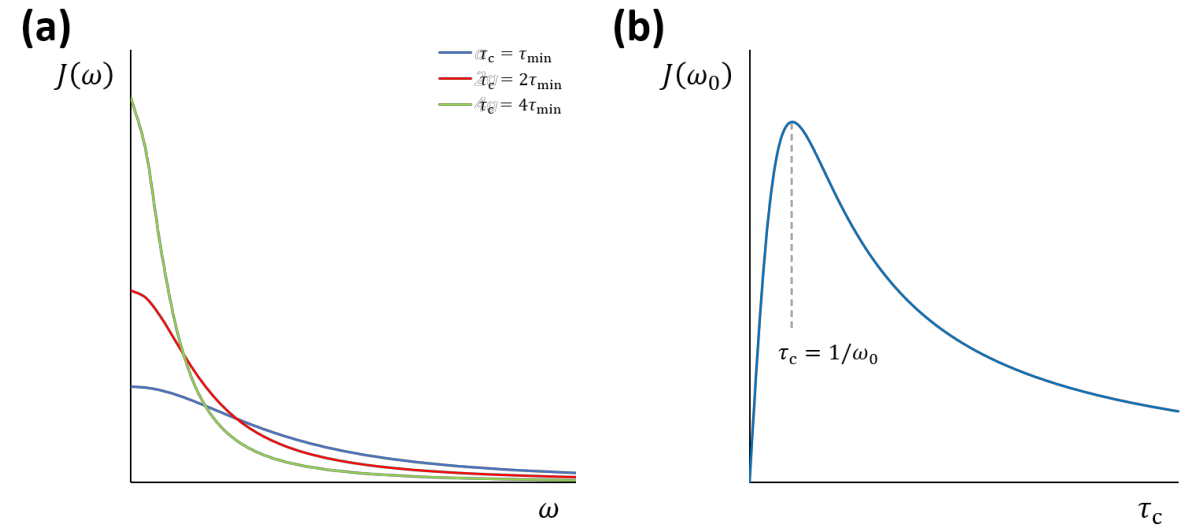


Figure 3.5: Spectral densities
(a) Plots of the spectral density function, *J*(ω), for different values of the correlation time, $\tau_c = \tau_{min}$, $2\tau_{min}$ and $4\tau_{min}$, showing that the shorter τ_c becomes, the higher the frequencies to which $J(\omega)$ spreads. **(b)** A plot of the spectral density at the Larmor frequency, $J(\omega_0)$, as a function of the correlation time, τ_c, indicating that the rate of longitudinal relaxation reaches a maximum at $\tau_c = 1/\omega_0$.

For simplicity, it is conventional to introduce the reduced spectral density, $j(\omega)$:

$$J(\omega) = \overline{B_{loc}^2} j(\omega) \tag{3.21}$$

$$j(\omega) = \frac{2\tau_c}{1+\omega^2\tau_c^2} \tag{3.22}$$

In the fast motion regime, where $\omega_0\tau_c \ll 1$, such as would be the case for small molecules, the reduced spectral density at the Larmor frequency, $j(\omega_0)$, is given by:

$$j(\omega_0) = \frac{2\tau_c}{1+\omega_0^2\tau_c^2} \approx 2\tau_c \tag{3.23}$$

This indicates that in the fast motion limit, the spectral density is independent of the Larmor frequency.

In the slow motion regime, such as would be the case for proteins, $\omega_0\tau_c \gg 1$, $(1 + \omega_0^2\tau_c^2) \approx \omega_0^2\tau_c^2$, and since $j(0) = 2\tau_c$, hence

$$j(\omega_0) \approx \frac{2}{\omega_0^2\tau_c} = \frac{j(0)}{\omega_0^2\tau_c^2} \tag{3.24}$$

As $\omega_0\tau_c \gg 1$, this means $j(\omega_0) \ll j(0)$.

The longitudinal relaxation rate constant, R_z or R_1, is given by:

$$R_1 = \gamma^2 J(\omega_0) \tag{3.25}$$

R_1 is typically estimated by an inversion-recovery experiment (Figure 3.6a):

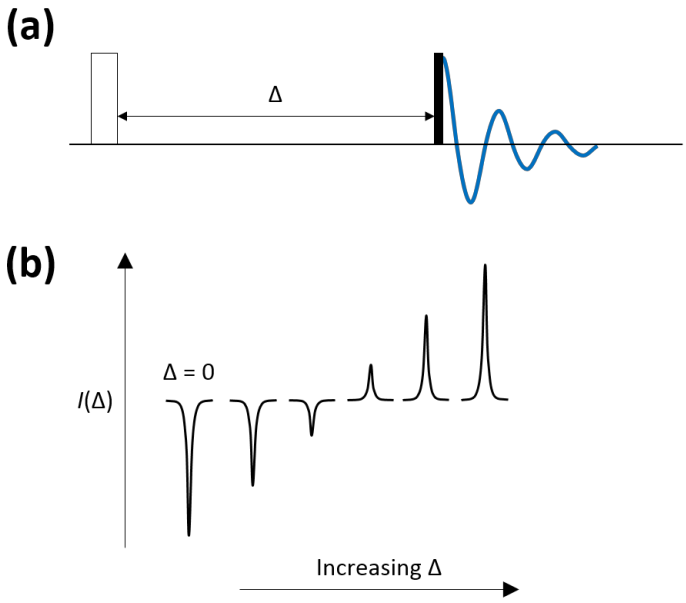


Figure 3.6: Pulse sequence (a) and typical set of spectra (b) for the inversion-recovery experiment
90° pulses are represented by black lines and 180° pulses are represented as white boxes. Pulses are applied with phase *x* unless indicated otherwise. The spectrum recorded at $\Delta = 0$ is phased so that the peak is negative, and the same phase correction has been applied to all other spectra. The peak intensities, $I(\Delta)$, are measured as a function of Δ.

The magnetisation is first inverted by a 180° pulse. The inverted magnetisation is then allowed to relax for a time Δ, after which a 90° pulse is applied to enable FID detection in the transverse plane. The peak intensity, $I(\Delta)$, as a function of longitudinal relaxation time Δ is:

$$I(\Delta) = I_0(2e^{-R_1\Delta} - 1) \tag{3.26}$$

where I_0 is the initial peak height, and R_1 the longitudinal relaxation rate.

Spectra are recorded with increasing Δ, and R_1 is estimated from peak heights measured as a function of Δ (Figure 3.6b).

There are two contributions to transverse relaxation: secular and non-secular. The secular contribution arises from there being a distribution of local fields along the z-axis with different Larmor frequencies, and depends on the spectral density at zero frequency, $J(0)$. The non-secular contribution involves reorientation of the magnetic moments of individual spins and hence, similar to longitudinal relaxation, depends on the spectral density at the Larmor frequency, $J(\omega_0)$. The rate constant for transverse relaxation, R_{xy} or R_2, is given as:

$$R_2 = \underbrace{\frac{1}{2}\gamma^2 J(0)}_{\text{secular}} + \underbrace{\frac{1}{2}\gamma^2 J(\omega_0)}_{\text{non-secular}} \tag{3.27}$$

The non-secular contribution is precisely half the value of the longitudinal rate constant, suggesting that the field is half as effective at causing transverse relaxation as it is at causing longitudinal relaxation.

R_2 can be estimated by measuring the width of the peak at half height, as Fourier transformation of the FID yields a peak with width at half-height of $2R_2$ rad s^{-1} or R_2/π Hz. However, R_2 can only be reliably estimated from linewidths if the line is homogeneously broadened. Such broadening is a fundamental property of the molecule, its environment and motion. Another contribution to rapid relaxation (broad linewidths) is inhomogeneous broadening, which refers to different parts of the sample precessing at slightly different Larmor frequencies (e.g. due to slight differences in magnetic field strength across the sample volume), and the increasing mismatch in magnetisation orientations across the sample over time leads to cancellation and decay of the total magnetisation.

To circumvent the problem of inhomogeneous contribution, R_2 is typically estimated using the Carr-Purcell-Meiboom-Gill (CPMG) relaxation experiment, where the decay of transverse magnetisation is monitored in a series of spin-echo pulse sequence elements. Any effects of inhomogeneous broadening would be refocused at the end of each spin-echo, whereas effects of homogeneous broadening cannot be refocused by spin-echos. Thus, the size of transverse magnetisation only depends on transverse relaxation which has taken place during the experiment.

R_1 and R_2 have different values in different motional regimes. In the fast motion limit, $j(\omega_0) = j(0) = 2\tau_c$:

$$R_1 = \gamma^2\overline{B_{loc}^2}j(\omega_0) = 2\gamma^2\overline{B_{loc}^2}\tau_c \tag{3.28}$$

$$\begin{aligned} R_2 &= \frac{1}{2}\gamma^2\overline{B_{loc}^2}j(0) + \frac{1}{2}\gamma^2\overline{B_{loc}^2}j(\omega_0) \\ &= \gamma^2\overline{B_{loc}^2}\tau_c + \gamma^2\overline{B_{loc}^2}\tau_c \\ &= 2\gamma^2\overline{B_{loc}^2}\tau_c \end{aligned} \tag{3.29}$$

Therefore, in the fast motion limit (e.g. for small molecules), R_1 and R_2 are equal.

In the slow motion limit, $j(\omega_0) \approx \frac{2}{\omega_0^2\tau_c}$ and $j(0) = 2\tau_c$, where $j(\omega_0) \ll j(0)$:

$$R_1 = \gamma^2\overline{B_{loc}^2}j(\omega_0) = \frac{2\gamma^2\overline{B_{loc}^2}}{\omega_0^2\tau_c} \tag{3.30}$$

$$R_2 = \tfrac{1}{2}\gamma^2\overline{B_{loc}^2}j(0) + \tfrac{1}{2}\gamma^2\overline{B_{loc}^2}j(\omega_0) = \gamma^2\overline{B_{loc}^2}\tau_c \tag{3.31}$$

Hence, as τ_c increases, R_1 reaches a maximum at $\tau_c = 1/\omega_0$. As τ_c increases further, R_1 decreases steadily. In contrast, R_2 continues to increase steadily with τ_c (Figure 3.7). For large molecules (e.g. proteins) which tumble slowly and have long τ_c, $R_2 > R_1$. The practical consequence of these observations is that transverse magnetisation decays to zero much more quickly than the *z*-magnetisation recovers to equilibrium.

^{15}N R_1 depends on $J(\omega_N)$, $J(\omega_H - \omega_N)$ and $J(\omega_H + \omega_N)$:

$$R_1 = \left(\frac{d^2}{10}\right)\{3J(\omega_N) + J(\omega_H - \omega_N) + 6J(\omega_H + \omega_N)\} + c^2J(\omega_N) \tag{3.32}$$

^{15}N R_2 depends on $J(0)$, $J(\omega_H)$, $J(\omega_N)$, $J(\omega_H - \omega_N)$ and $J(\omega_H + \omega_N)$:

$$R_2 = \left(\frac{d^2}{20}\right)\{4J(0) + 3J(\omega_N) + J(\omega_H - \omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)\} + \left(\frac{c^2}{6}\right)\{4J(0) + 3J(\omega_N)\} + R_{ex} \tag{3.33}$$

where $d^2 = \left(\frac{\gamma_N\gamma_H\hbar}{r_{NH}^3}\right)^2$, $c^2 = \frac{2}{15}[\gamma_N B_0(\sigma_\parallel - \sigma_\perp)]^2$, and are coefficients for dipolar and CSA interactions, respectively, and R_{ex} is the chemical exchange contribution to R_2 value [324].

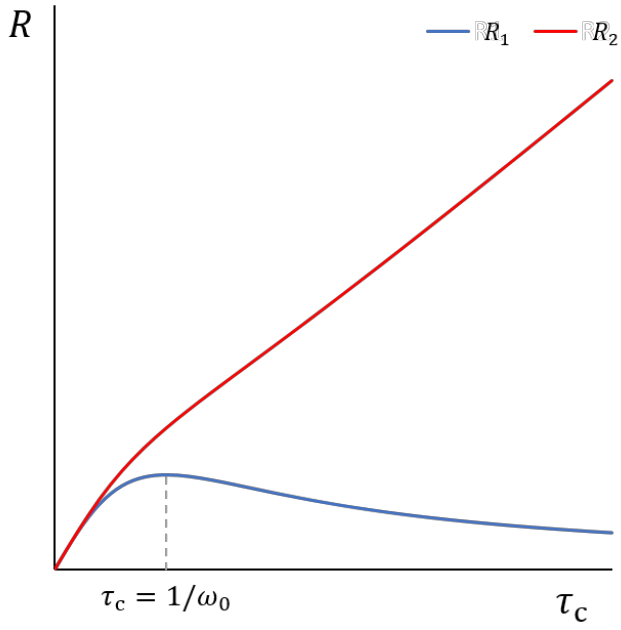


Figure 3.7: Variations in R_1 and R_2 with overall rotational correlation time, τ_c

There are various mathematical models for mapping the spectral density functions, of which the Lipari-Szabo model-free analysis [325,326] is widely used to obtain information about site-specific internal motions of proteins. However, without any assumptions, the spectral density functions at five frequencies cannot be determined from experimentally determined relaxation rate constants. According to the model-free formalism of Lipari and Szabo, the global motion of the whole molecule and internal motions of ^{15}N-^{1}H bond vectors are assumed to be independent. Internal motion is defined by two parameters: the square of the order parameter, S^2, for describing the amplitude, and a characteristic correlation time of the motion τ_i.

$J(\omega)$ depends on both the overall motion of the macromolecule as a whole and on the internal motions of the ^{15}N-^{1}H bond vector. The formalism enables simplification of the spectral density function:

$$J(\omega) = \frac{2}{5}\left(\frac{S^2\tau_c}{1+\omega^2\tau_c^2} + \frac{(1-S^2)\tau_e}{1+\omega^2\tau_e^2}\right) \quad (3.34)$$

where τ_e is the effective internal correlation time and τ_c is the isotropic global correlation time, and are defined by the following relationship:

$$\frac{1}{\tau_e} = \frac{1}{\tau_c} + \frac{1}{\tau_i} \quad (3.35)$$

In the limit $\tau_e \ll \tau_c$ (effective internal correlation time is much smaller than the isotropic global correlation time) and $S^2 \rightarrow 1$ (relaxation dominated by global motion), the simplified spectral density function is:

$$J(\omega) = \frac{2}{5}\frac{S^2\tau_c}{(1+\omega^2\tau_c^2)} \tag{3.36}$$

Under this limit, expressions for ^{15}N R_1 and ^{15}N R_2 can be approximated as follows:

$$R_1 \propto J(\omega_N) \cong \frac{S^2}{\omega_N^2\tau_c} \tag{3.37}$$

$$R_2 \propto J(0) \cong S^2\tau_c \tag{3.38}$$

R_1 is sensitive to the dynamics on the timescale of ps–μs, whereas R_2 is sensitive to motions on both the ps–μs and μs–ms timescales.

The above assumption is also made in the R_2/R_1 ratio method, so that R_2/R_1 is essentially independent of S^2 and τ_e:

$$R_2/R_1 \propto \omega_N^2\tau_c^2 \tag{3.39}$$

Using R_2/R_1 ratios which fall within one standard deviation of the mean, τ_c can be estimated as [327]:

$$\tau_c = \frac{1}{4\pi\omega_N} \times \sqrt{\frac{6R_2}{R_1} - 7} \tag{3.40}$$

Accurate assessment and interpretation of spin-relaxation data requires careful consideration of anisotropic rotational diffusion and chemical exchange. While R_2/R_1 is a commonly used approach for identifying residues undergoing chemical exchange and estimation of global correlation times, this approach does not distinguish between motional anisotropy and chemical exchange. This leads to error due to potential contribution from R_{ex} terms to the R_2 relaxation rate, since $R_2 = {R_2}^0 + R_{ex}$. Kneller *et al.* recommend the use of the R_1R_2 value, which significantly attenuates the effects of motional anisotropy [328], as anisotropic tumbling has opposing effects on R_1 and R_2.

$$R_1R_2 \propto J(0)J(\omega_N) \cong \frac{S^4}{\omega_N^2} \tag{3.41}$$

3.2.5.3 Spin relaxation in the rotating frame ($R_{1\rho}$)

Spin relaxation in the rotating frame ($R_{1\rho}$) is another method for measuring transverse relaxation. In the $R_{1\rho}$ experiment, an RF field (B_1) is applied to spin-lock the magnetisation along the direction of the effective field in the rotating frame. This can be performed using near-resonance or off-resonance RF fields, and are called on-resonance (Figure 3.8b) or off-resonance (Figure 3.8a) $R_{1\rho}$ experiments, respectively. The result is that there is relaxation but no precession. Figure 3.8c illustrates the geometrical relationship between the laboratory and tilted rotating frames. $R_{1\rho}$ is the relaxation rate constant for magnetisation which is spin-locked along the direction of the effective field in the rotating frame using a spin-lock with field strength ω_1. $R_{1\rho}$ contains contributions from R_1 and R_2 due to transformation from the laboratory frame to the rotating frame [329]:

$$R_{1\rho} = R_1 \cos^2\theta + R_2 \sin^2\theta \tag{3.42}$$

where the tilt angle in the rotating frame is

$$\theta = \tan^{-1}\left(\frac{\omega_1}{\bar{\Omega}}\right) \tag{3.43}$$

the average resonance offset is

$$\bar{\Omega} = \sum_{i=1}^{N} p_i \Omega_i \tag{3.44}$$

and the average effective field in the rotating frame is

$$\omega_e = \sqrt{(\bar{\Omega}^2 + {\omega_1}^2)} \tag{3.45}$$

To obtain accurate R_2 from $R_{1\rho}$, it is advantageous to use a strong B_1 field strength, thereby making θ close to 90°.

The accessible range of effective magnetic field strengths determines the timescale of processes that can be studied by different relaxation experiments. The effective field strengths typically employed in CPMG relaxation experiments are 25–500 Hz; hence CPMG experiments are often used for characterising slower ms timescale chemical exchange processes. Chemical exchange kinetics are obtained from the variation in R_2 as a function of the time delay between refocusing pulses in the spin-echo sequence. The effective field strengths employed in $R_{1\rho}$ experiments are 1–6 kHz, although weaker fields can be utilised to provide overlap with the CPMG experiment. Thus, $R_{1\rho}$ experiments are most often used for

characterising faster µs timescale chemical exchange processes. Chemical exchange kinetics are obtained from the variation of $R_{1\rho}$ as a function of the effective field in the rotating frame.

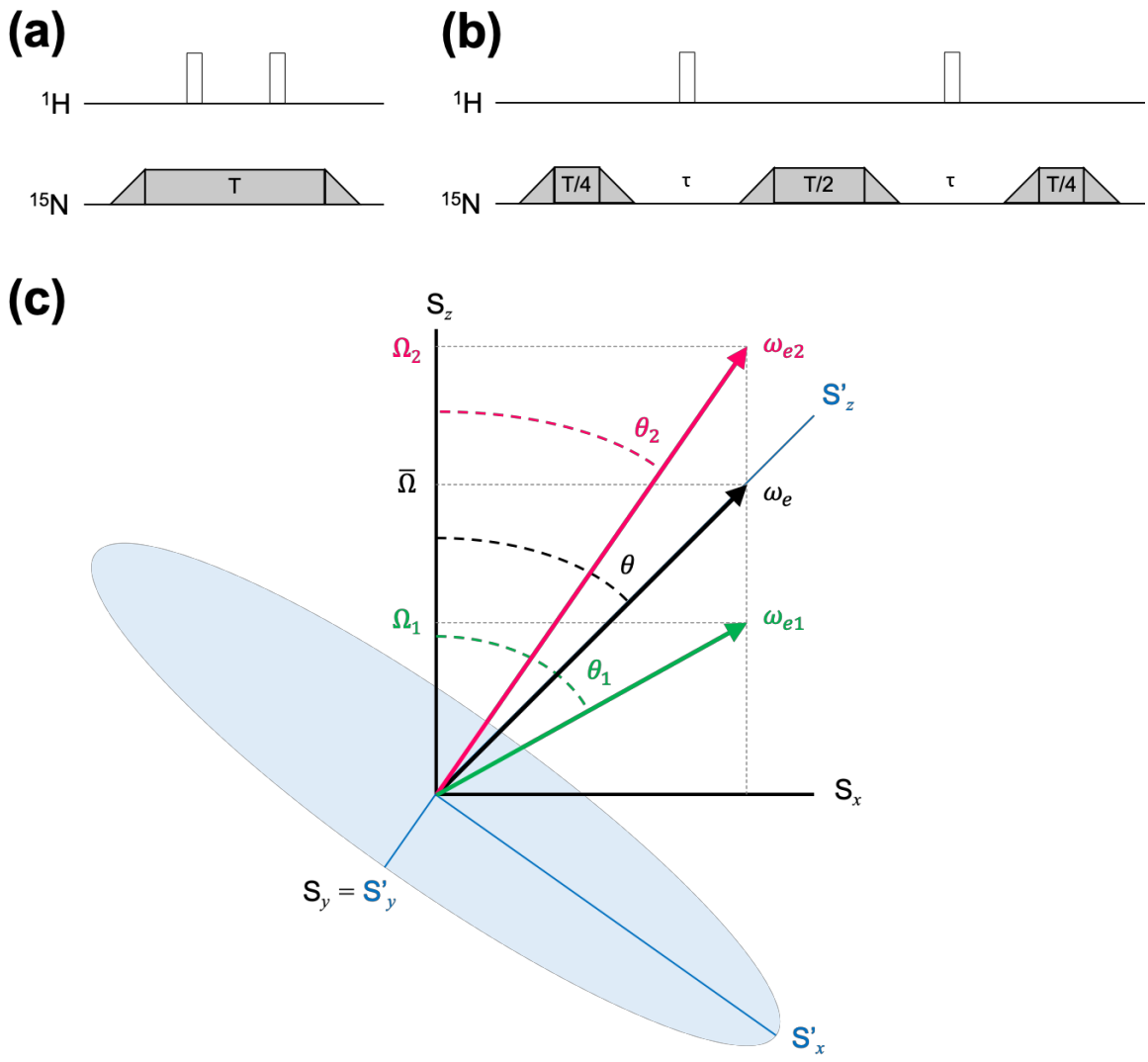


Figure 3.8: Basic pulse sequence elements for the measurement of $R_{1\rho}$ and geometric representation of the laboratory and rotating frames
Basic pulse sequence elements for off-resonance **(a)** and on-resonance **(b)** ^{15}N $R_{1\rho}$ measurements. Wide white bars indicate 180° pulses. Pulses are applied with phase *x* unless indicated otherwise. Grey rectangles represent ^{15}N spin-lock periods, which are applied with phase *x* for a total length of T. In both the off-resonance and on-resonance experiments, adiabatic sweeps (grey triangles) are used to rotate the magnetisation from the *z*-axis to the direction of the effective field, and vice versa. **(c)** Geometric representation of the laboratory frame (S_x, S_y, S_z) and the tilted rotating frame (S'_x, S'_y, S'_z), where S'_z is parallel to the average effective field, ω_e. The relative orientation of the laboratory and rotating frames is given by the tilt angle θ. The rotating frame and θ are different for each spin. The orientations of the effective fields of two different spins, ω_{e1} and ω_{e2}, are given by θ_1 and θ_2, respectively. Figure adapted from [329].

An advantage of the $R_{1\rho}$ experiment is that off-resonance effects can be corrected using Equation 3.42, provided that R_1 data is available. The accuracy of $R_{1\rho}$ measurements decreases for signals located far off-resonance. Compared to CPMG R_2, $R_{1\rho}$ values do not need to be recorded at different carrier frequencies without discarding any data.

In CPMG R_2, off-resonance error is significant at high magnetic field strength. The CPMG pulse train accumulates error caused by a combination of pulse imperfections and off-resonance effects. The main disadvantage is that there is no simple equation to correct for CPMG R_2 off-resonance effects. In practice, it is recommended to discard R_2 data obtained at $n/2\tau_{CP}$ frequency (τ_{CP} is half the duration of CPMG interpulse delay and n is an integer), and to record the data at two different carrier frequencies [330]. Alternatively, phase cycling can be used to average out off-resonance effects [199].

3.2.5.4 Other relaxation parameters

In addition to the R_1 and R_2 relaxation rate constants, relaxation behaviour in two-spin systems introduces additional features which are not present in one-spin systems, particularly cross relaxation which gives rise to nuclear Overhauser effect (NOE). Relaxation via the NOE occurs between two spins which are interacting via DD coupling. Each spin is described by two energy levels corresponding to the α state and the β state. DD coupling leads to relaxation-induced transitions between any of the four levels. The rate constant σ_{12} describes the rate at which magnetisation is transferred from spin 1 by relaxation processes to spin 2:

$$\sigma_{12} = \frac{b^2}{20}\left[6J\left(\omega_{0,1} + \omega_{0,2}\right) - J\left(\omega_{0,1} - \omega_{0,2}\right)\right] \tag{3.46}$$

where $b = \frac{\mu_0 \gamma_1 \gamma_2 \hbar}{4\pi r^3}$, and $\omega_{i,j}$ is the frequency needed to cause transitions between two levels i and j.

For slow tumblers, σ_{12} is negative, as $6J\left(\omega_{0,1} + \omega_{0,2}\right) < J\left(\omega_{0,1} - \omega_{0,2}\right)$.
For rapid tumblers, σ_{12} is positive, as $6J\left(\omega_{0,1} + \omega_{0,2}\right) > J\left(\omega_{0,1} - \omega_{0,2}\right)$.

σ_{12} is inversely proportional to r^6, where r is the distance between the two spins. Hence, σ_{12} falls off very rapidly with distance, and is detectable for inter-spin distances within 5 Å for the case of two protons.

Heteronuclear NOE experiments were not pursued in this chapter, as experimental results on pSRII (see Section 3.3.6) and published studies on SDS-solubilised proteolytic fragment of bacteriorhodopsin [331] indicate that the timescales of NH bond vector motions and chemical exchange are not within the ps–ns timescale, for which heteronuclear NOE experiments are most suited. Moreover, the signal-to-noise ratio is low for big proteins and artifacts are common due to cross correlation when measured at high static magnetic field strengths, which are required to provide sufficient spectral dispersion for unfolded proteins.

3.3 Results: SDS at pH 6.0

In Chapter 2, denaturant screening and extensive biophysical characterisation identified SDS at pH 6.0 and pH 2.0 as the most suitable conditions for unfolding pSRII. Chemically denatured states constitute an ensemble of rapidly interconverting conformations which are present at the beginning of the folding pathway. NMR is therefore highly suited for yielding atomic resolution information on structure and dynamics represented by a heterogeneous ensemble of unfolded protein conformations.

3.3.1 SDS-denatured pSRII remains partially structured, with conformations in the denatured states ensemble interconverting on the µs–ms timescale

NMR spectra of pSRII in c7-DHPC at pH 6.0 are typically recorded at 35 °C or higher temperatures of up to 60 °C to yield good spectral resolution [280]. This is because higher temperatures lead to lower viscosity and a decrease in transverse relaxation rate, hence yielding narrower linewidths [332]. Due to the higher pSRII concentrations necessary for NMR samples, 0.5% c7-DHPC was included in NMR samples of unfolded pSRII to ensure the presence of at least 40-fold excess [333] of c7-DHPC detergent molecules.

Circular dichroism (Figure 3.9a) and tryptophan fluorescence (Figure 3.9b) were first recorded for pSRII in SDS + 0.5% c7-DHPC at 35 °C, pH 6.0 for comparable studies against biophysical measurements done at 25 °C (see Chapter 2). The extent of unfolding was assessed by changes in MRE at 222 nm and fluorescence intensity at 335 nm, showing that unfolding occurs to the same extent at 35 °C. This indicates that results from the biophysical studies in Chapter 2 would be directly comparable to pSRII unfolded in SDS at 35 °C for NMR experiments. The rate of aggregation appears to be faster at 35 °C, since decrease in α-helical content and oligomerisation are already observable by 6 h (Figures 3.9a & 3.9c).

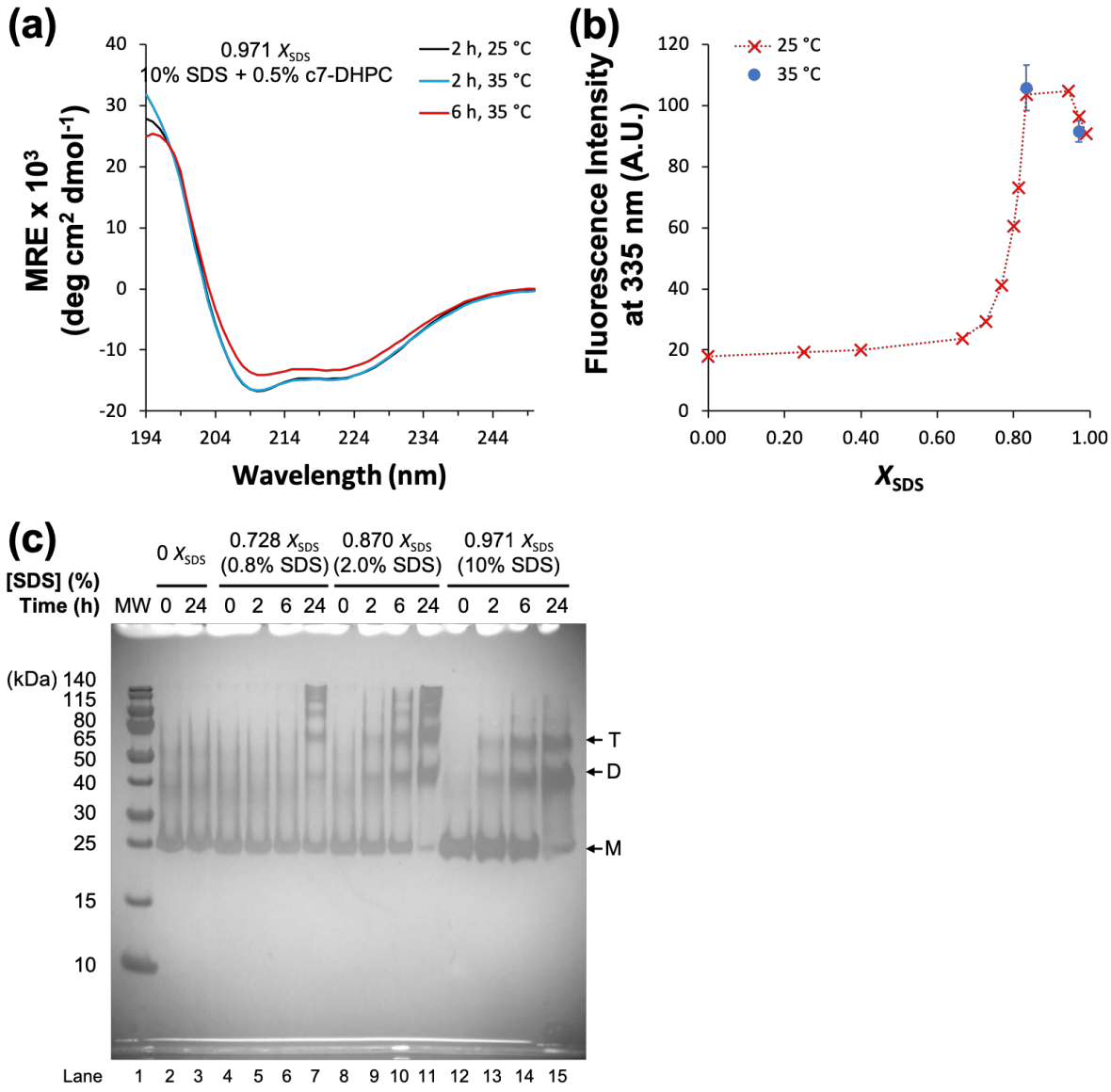


Figure 3.9: Circular dichroism (a), fluorescence emission (b) and SDS-PAGE (c) of pSRII in different SDS + 0.5% c7-DHPC at 35 °C
(a) MRE at 222 nm for 7 µM (0.2 mg/ml) pSRII in 0.971 X_{SDS} (10% SDS + 0.5% c7-DHPC) at 25 or 35 °C for up to 6 h. **(b)** Tryptophan fluorescence emission at 335 nm. Error bars represent the standard deviation of three independent measurements. **(c)** Silver-stained 12% SDS-PAGE showing increased aggregation of pSRII with increasing X_{SDS} over prolonged unfolding time (0 min, 2 h, 6 h, 24 h). Lane 1, molecular weight marker; lanes 2–3, 0 X_{SDS} (0.5% c7-DHPC); lanes 4–7, 0.728 X_{SDS} (0.8% SDS + 0.5% c7-DHPC); lanes 8–11, 0.870 X_{SDS} (2.0% SDS + 0.5% c7-DHPC); lanes 12–15, 0.971 X_{SDS} (10% SDS + 0.5% c7-DHPC). The arrows indicate the expected migration of monomeric (M), dimeric (D) and trimeric (T) pSRII.

1D ^{1}H NMR spectra were recorded to assess the structural properties of SDS-denatured pSRII (Figure 3.10). Folded pSRII (0 X_{SDS}) showed wide chemical shift dispersion from about 6 to 10 ppm. The dispersion remained equally wide in the presence of increasing X_{SDS},

indicating that many backbone amide residues still experience unique chemical environments. This supports the results from circular dichroism experiments (see Chapter 2) showing that large portions of helical structure were preserved across all X_{SDS} studied. An obvious drop in peak intensities was observed for pSRII in 0.834 X_{SDS} (1.5% SDS in 0.5% c7-DHPC), indicating substantial changes in protein backbone dynamics, either due to a slower tumbling rate in the presence of more detergent or due to conformational exchange in the μs–ms timescale (intermediate exchange regime). This coincides with the onset of full unfolding characterised by biophysical studies (see Chapter 2). Tryptophan side-chain indole groups also experience changes in dynamics, as evidenced by a similar drop in peak intensities in the 9.5–10.5 ppm region (Figure 3.10, left inset). Increased peak overlap was observed upon further increase in X_{SDS} up to 0.971 X_{SDS} (10% SDS + 0.5% c7-DHPC), suggesting that as the protein is further denatured more residues become exposed to similar chemical environments. Beyond 0.971 X_{SDS}, more residues undergo conformational exchange in the μs–ms timescale, as a decrease in peak intensities was observed in 0.990 X_{SDS} (30% SDS in 0.5% c7-DHPC). Similar observations were made for the 5.5–7.0 ppm region (Figure 3.10, right inset), part of which possibly shows changes in arginine side-chain amide proton signals.

To understand how spectral quality is affected by pSRII unfolding and aggregation, time-dependent changes in ^{1}H NMR spectra were studied for pSRII in 0.834 X_{SDS} (1.5% SDS in 0.5% c7-DHPC) (Figure 3.11a) and in 0.971 X_{SDS} (10% SDS in 0.5% c7-DHPC) (Figure 3.11b). Spectra were recorded at 10 min, 5.3 h and 21.3 h after initial exposure of pSRII to SDS, and the samples were kept at 35 °C throughout. Both samples showed complete retinal Schiff base hydrolysis and became pale yellow in colour (absorbance λ_{max} = 390 nm) after 21 h. Comparison of ^{1}H spectra recorded at 10 min and 5.3 h showed a very small decrease in peak intensities, with different signals showing different amounts of changes, indicating differential changes in protein backbone dynamics across the protein. A more pronounced decrease in peak intensities was observed between 5.3 h and 21.3 h, indicative of aggregation. In 0.834 X_{SDS} (1.5% SDS in 0.5% c7-DHPC), several peaks (indicated by arrows in Figure 3.11a) were noted to show higher intensities at 21.3 h than at the earlier time points, including a tryptophan side-chain indole signal at 10 ppm. This indicates that residue-specific changes in protein structure and dynamics might have occurred as well. Similar conclusions could be drawn from the time-dependent spectral changes in

0.971 X_{SDS} (10% SDS in 0.5% c7-DHPC), as peaks at 6.0–6.5 ppm have very similar intensities at 5.3 h and 21.3 h whereas other peaks show pronounced decrease in intensities.

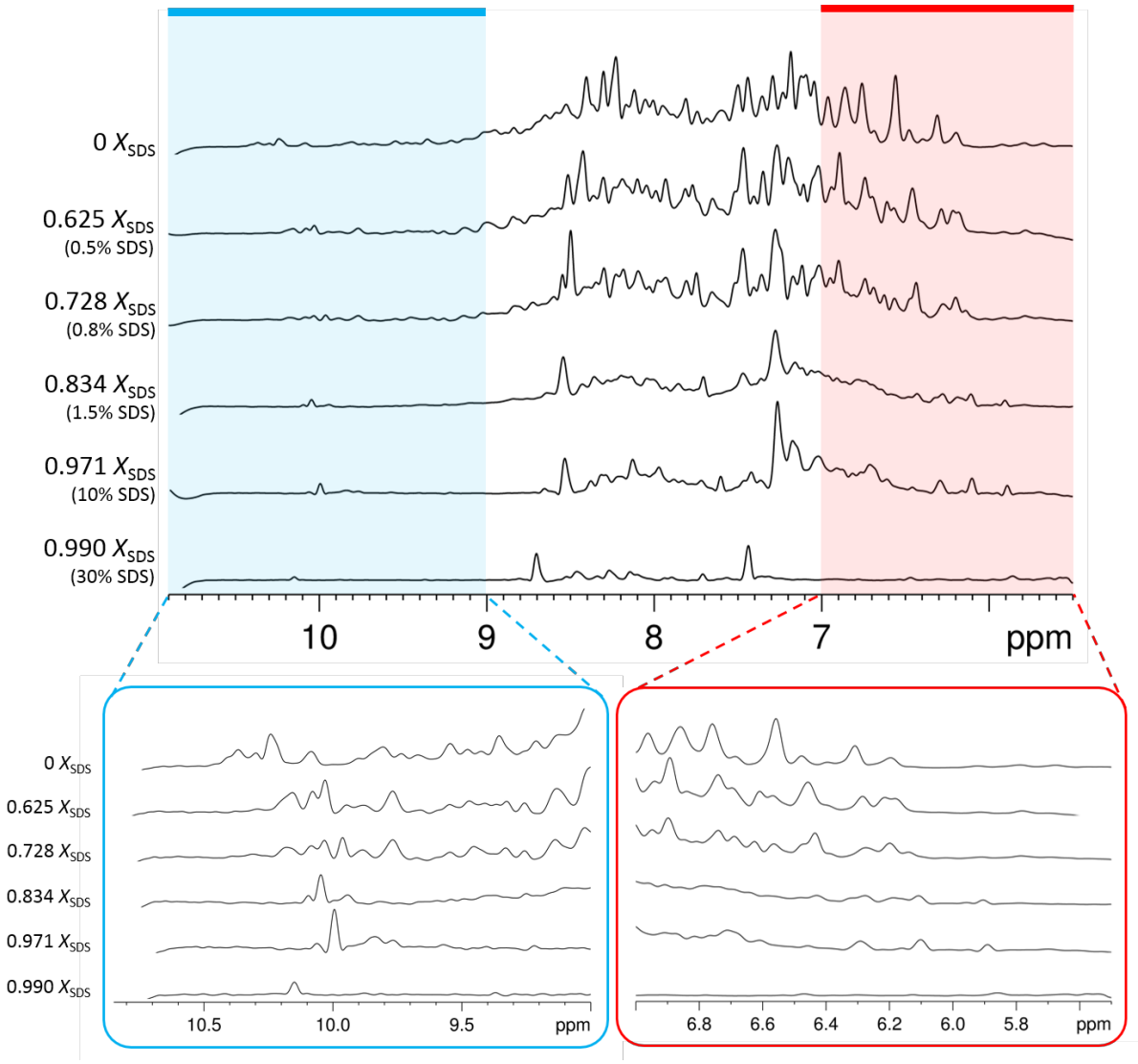


Figure 3.10: 1D ^{1}H NMR of pSRII in different X_{SDS}
All samples contain ~ 83 µM pSRII and 0.5% c7-DHPC. All spectra were recorded at 308 K, with 512 scans, at ~ 10 min after initial exposure to SDS. All spectra were processed by apodization of the FID using a squared sine window function shifted by 90° followed by an exponential line broadening of 1 Hz. Insets show expanded views for the 9.0–11.0 ppm region (left) and the 5.5–7.0 ppm region (right). All samples contained TSP as internal reference.

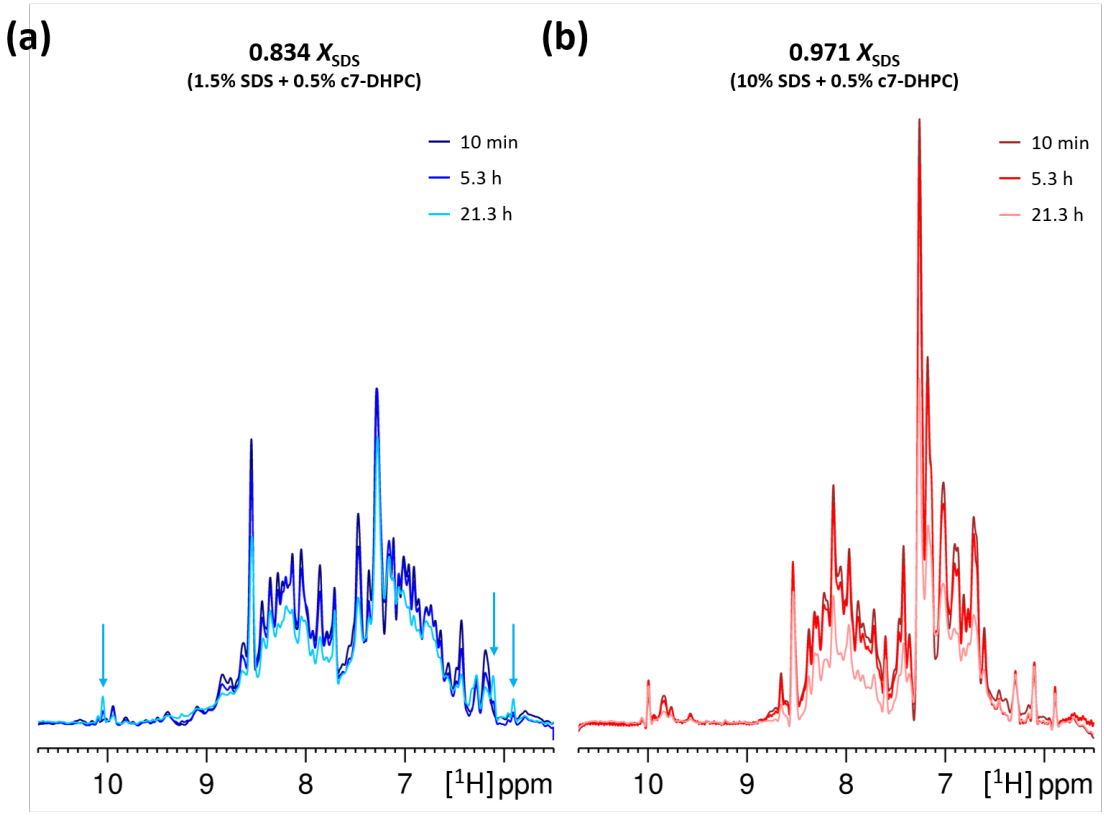


Figure 3.11: Changes in ^{1}H NMR spectra over time
^{1}H NMR of backbone amides and side-chain NH groups for 83 µM pSRII in 0.834 X_{SDS} (1.5% SDS in 0.5% c7-DHPC) **(a)** or 0.971 X_{SDS} (10% SDS in 0.5% c7-DHPC) **(b)** for 10 min, 5.3 h or 21.3 h. All spectra were recorded at 308 K with 512 scans, and processed by apodization of the FID using a squared sine window function shifted by 90° followed by an exponential line broadening of 1 Hz. All samples contained TSP as internal reference.

Residue-specific information on SDS-denatured pSRII could not be obtained from 1D ^{1}H NMR spectra due to the large number of overlapping backbone amide proton signals. The ^{19}F nucleus is highly sensitive (has a high gyromagnetic ratio comparable to that of ^{1}H), shows a broad chemical shift range, and does not occur naturally in biomacromolecules. Selective ^{19}F-labelling of specific amino acids would significantly decrease the number of overlapping signals and enable facile spectral interpretation.

^{19}F-tryptophan labelling of pSRII was accomplished using the method outlined by Crowley *et al* [334]. As fluorine is not identical to the replaced proton, structural perturbations were assessed by circular dichroism and 2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC. No secondary structure changes were detected in the absence of SDS (Figure 3.12a), but ^{19}F-5-Trp labelled

pSRII was more α-helical than unlabelled pSRII in the presence of 0.943 X_{SDS} (1.0% SDS in 0.1% c7-DHPC) (Figure 3.12b), indicating that ^{19}F-labelling is not completely inert. Previous reports suggest that fluorinated amino acid incorporation could affect protein stability through increased hydrophobicity and even alter the pKa of exchangeable protons [335].

From the 2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC spectra (Figure 3.12c) of uniformly ^{15}N-labelled pSRII and ^{15}N,^{19}F-5-Trp-labelled pSRII, the combined chemical shift difference was calculated as the square root of the scaled sum of squares of chemical shift changes in the ^{1}H and ^{15}N dimensions:

$$\Delta\delta(^{1}\mathrm{H},^{15}\mathrm{N}) = \sqrt{\Delta\delta(^{1}\mathrm{H})^{2} + \frac{\Delta\delta(^{15}\mathrm{N})^{2}}{6}} \quad (3.47)$$

Most backbone amides experience $\Delta\delta(^{1}\mathrm{H},^{15}\mathrm{N})$ smaller than 0.1 ppm, indicating that ^{19}F-5-Trp labelling does not affect the overall protein structure. Residues with $\Delta\delta(^{1}\mathrm{H},^{15}\mathrm{N}) \geq 0.1$ ppm are mainly localised near the sites of fluorine incorporation (Figure 3.12d). Interestingly, tryptophan side-chain indoles were unobservable, suggesting that fluorine incorporation might have led to changes in the relaxation properties of the side-chain indole groups.

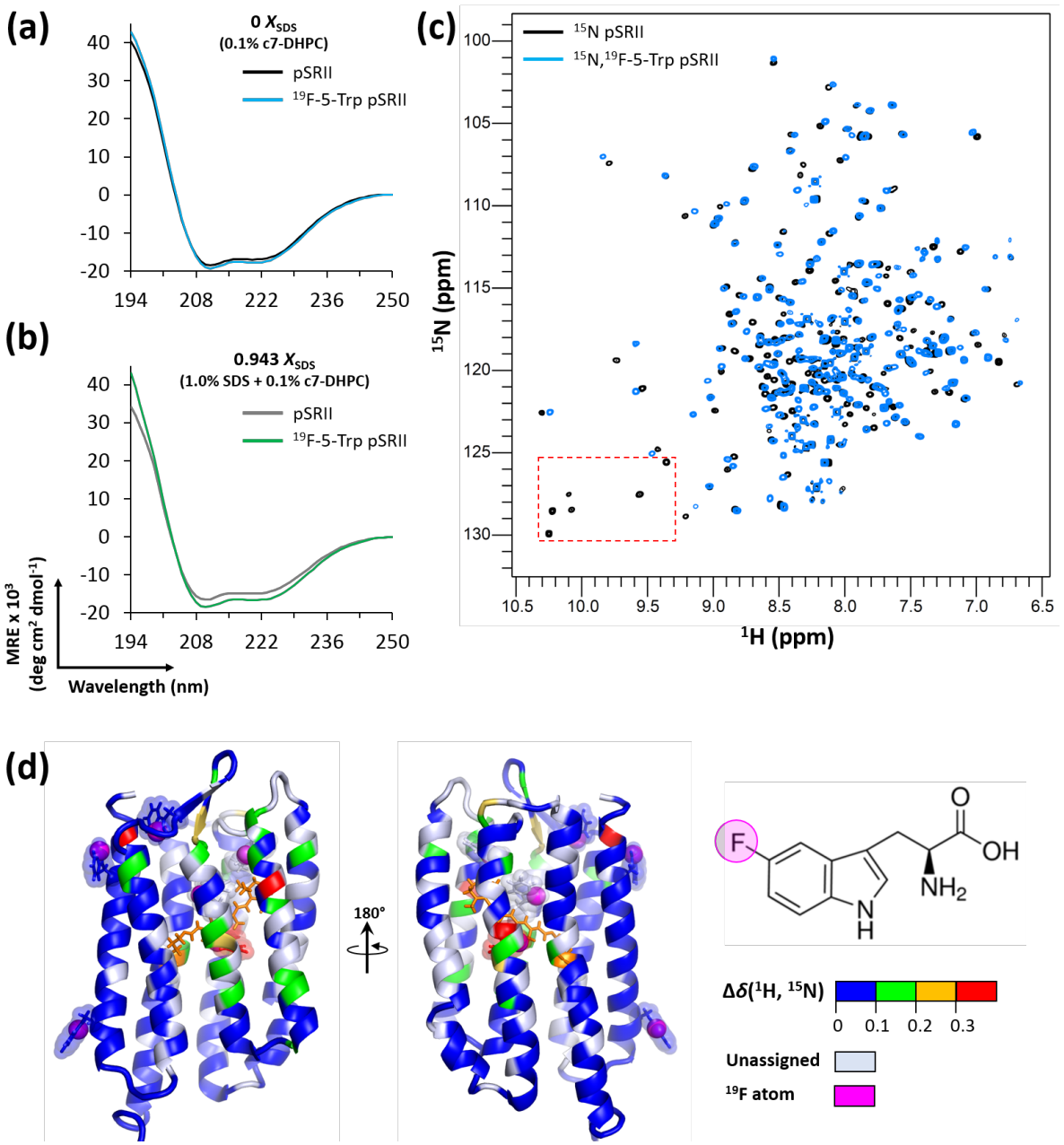


Figure 3.12: ^{19}F-5-Trp labelling of pSRII leads to minor changes in protein structure
(a–b) Circular dichroism. **(c)** [^{1}H,^{15}N] SOFAST-TROSY-HSQC (80 scans, 308 K) of 83 µM ^{15}N-pSRII (black) in 0.5% c7-DHPC and 157 µM ^{15}N,^{19}F-5-Trp-pSRII in 1.8% c7-DHPC (blue), showing good spectral overlap indicative of limited structural changes induced by ^{19}F-labelling. Trp side-chain indoles have become unobservable (red dashed box). Spectra were recorded with TSP as internal reference. **(d)** Combined chemical shift differences, Δδ(^{1}H,^{15}N), are small across most backbone amides, except for selected residues which are located close to the sites of fluorine incorporation.

A 1D ^{19}F NMR spectrum of pSRII in 0 X_{SDS} shows six well-resolved peaks corresponding to the 6 tryptophan residues (Figure 3.13a). In the presence of increasing X_{SDS} up to 0.834 X_{SDS} (1.5% SDS in 0.5% c7-DHPC), increased peak overlap and line broadening were observed (Figure 3.13a), indicating that some tryptophan residues became exposed to similar chemical environments while slow-to-intermediate conformational exchange appeared to affect all tryptophan residues, consistent with the observations from 1D ^{1}H NMR. Only one resonance corresponding to the sum of peaks 5 and 6 could be observed in 0.870 X_{SDS} (2.0% SDS in 0.5% c7-DHPC), suggesting that the tryptophan residues are exposed to similar chemical environments. The chemical environments of tryptophans corresponding to peaks 5 and 6 became different again in 0.971 X_{SDS} (10% SDS in 0.5% c7-DHPC), manifesting as a small shoulder peak. This could be correlated with the formation of non-native helical structures due to the helix-inducing effect of SDS (see Chapter 2). Assignment of the ^{19}F NMR signals was not pursued given the low information content of this approach due to substantial peak overlaps. Comparison of ^{19}F spectra recorded at 1 h, 4 h and 23 h showed more pronounced decrease in peak intensities in 0.834 X_{SDS} (1.5% SDS in 0.5% c7-DHPC) than in 0.400 X_{SDS} (0.2% SDS in 0.5% c7-DHPC) (Figure 3.13b), consistent with time-dependent aggregation of denatured pSRII beyond the unfolding transition.

In summary, only a limited amount of residue-specific information could be derived from 1D ^{1}H and ^{19}F NMR spectra due to:

1) a large number of backbone amide proton signals in 1D ^{1}H NMR spectra across all X_{SDS},
2) peak overlap in 1D ^{19}F NMR spectra in high X_{SDS} as the tryptophan residues were exposed to similar chemical environments upon protein unfolding, and
3) line broadening which could be attributed to conformational exchange on the μs–ms timescale and/or aggregation at ≥ 6 h after initial exposure to SDS.

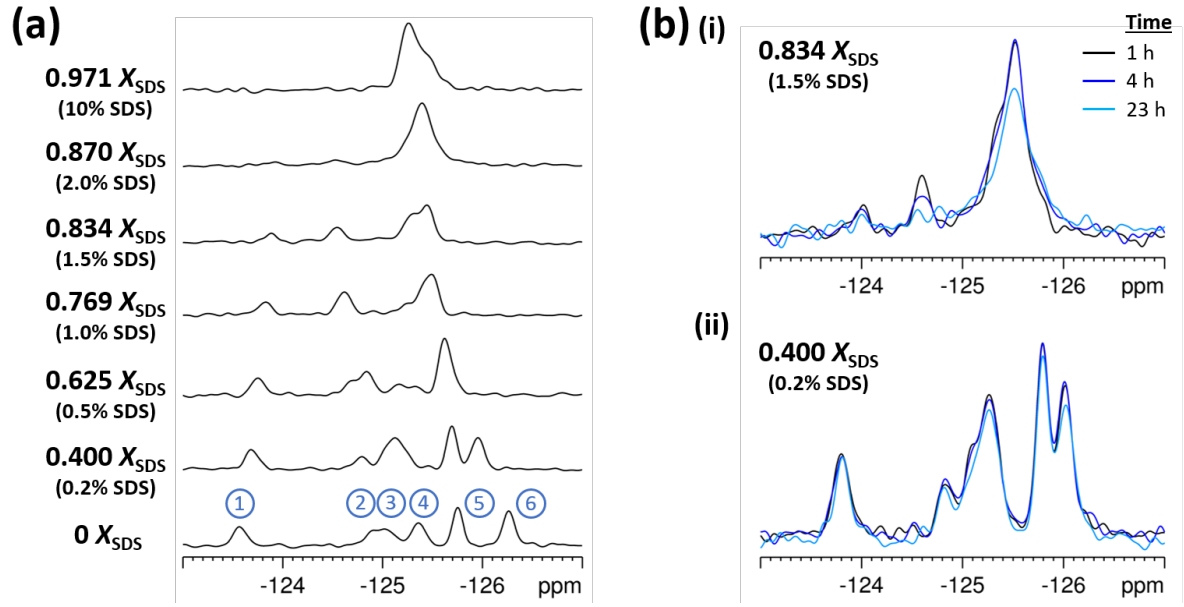


Figure 3.13: 1D ^{19}F NMR of ^{19}F-5-Trp pSRII in different X_{SDS} (a) and over time (b)
All spectra were recorded at 308 K, at ~ 10 min after initial exposure to SDS, with 512 scans. All spectra were processed by apodization of the FID using a squared sine window function shifted by 90° followed by an exponential line broadening of 1 Hz. All spectra were referenced externally to $CFCl_3$ (0 ppm).

3.3.2 Chemical shift: Structural properties of SDS-denatured pSRII

Given the spectral overlap observed for backbone amide protons in 1D ^{1}H NMR and at high X_{SDS} in 1D ^{19}F NMR, 2D NMR will likely be more suitable for resolving overlapped signals to obtain residue-specific information on SDS-unfolded pSRII.

2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC experiments were recorded to investigate the effects of SDS-mediated unfolding on the backbone amides of pSRII. Assignments for backbone amide resonances and the solution-state NMR structure of pSRII in 0 X_{SDS} at pH 6 were available from previous work in the Nietlispach lab [216].

Figure 3.14 shows an overlay of [^{1}H,^{15}N] SOFAST-TROSY-HSQC spectra recorded in the presence of 0–0.990 X_{SDS} (0–30% SDS in 0.5% c7-DHPC). Increasing X_{SDS} led to progressive changes in chemical shifts (see arrows in Figure 3.14b). Peak intensities generally decreased with increasing X_{SDS}, as evidenced by most residues being no longer observable in ≥ 0.870 X_{SDS} (Figure 3.14b). The overall decrease in peak intensities with

increasing X_{SDS} is attributed predominantly to intermediate (μs–ms timescale) exchange between different backbone amide conformations represented by the ensemble of denatured states, and also to minor amounts of aggregation (Figure 3.9c). Residues in the C-terminal tail of pSRII remained significantly more flexible than the rest of the protein, as evidenced by the strong intensities of these peaks across all X_{SDS}. Big differences in peak intensities between transmembrane vs. C-terminal residues also indicate that the transmembrane helices are likely to remain embedded in SDS/c7-DHPC mixed micelles across all X_{SDS} and that there is little/no stable interaction of the C-terminal tail with detergent micelles.

To understand whether backbone amide chemical shifts are determined by X_{SDS} or the molar concentration of SDS, 2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC spectra of pSRII were recorded in 1% SDS + 0.45% c7-DHPC and in 2% SDS + 0.90% c7-DHPC, both corresponding to 0.769 X_{SDS} but with different molar concentrations of c7-DHPC and SDS (Figure 3.15). The peaks in both spectra have almost identical chemical shifts, indicating that chemical shifts are determined by X_{SDS}. This observation justifies correlating results from the biophysics studies in Chapter 2 against X_{SDS}, as this parameter is representative of the immediate micellar environment experienced by the protein. It is noted that peak intensities are slightly lower in 2% SDS + 0.90% c7-DHPC because the presence of higher detergent concentrations leads to higher viscosity and slower molecular tumbling, thus yielding slightly broader linewidths.

To understand how spectral quality is affected by pSRII unfolding and aggregation at 35 °C, time-dependent changes were studied in 2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC spectra of pSRII in 0.834 X_{SDS} (1.5% SDS in 0.5% c7-DHPC) (Figure 3.15b), which is beyond the unfolding transition. Comparison of the TROSY spectra over time showed very small decrease in peak intensities within 6 h of initial exposure to SDS, while more pronounced decrease in peak intensities was observed from 6 h onwards, indicative of aggregation. After being kept at 35 °C for 21 h, the sample showed complete retinal Schiff base hydrolysis (clear but pale yellow in colour, with absorbance λ_{max} = 390 nm).

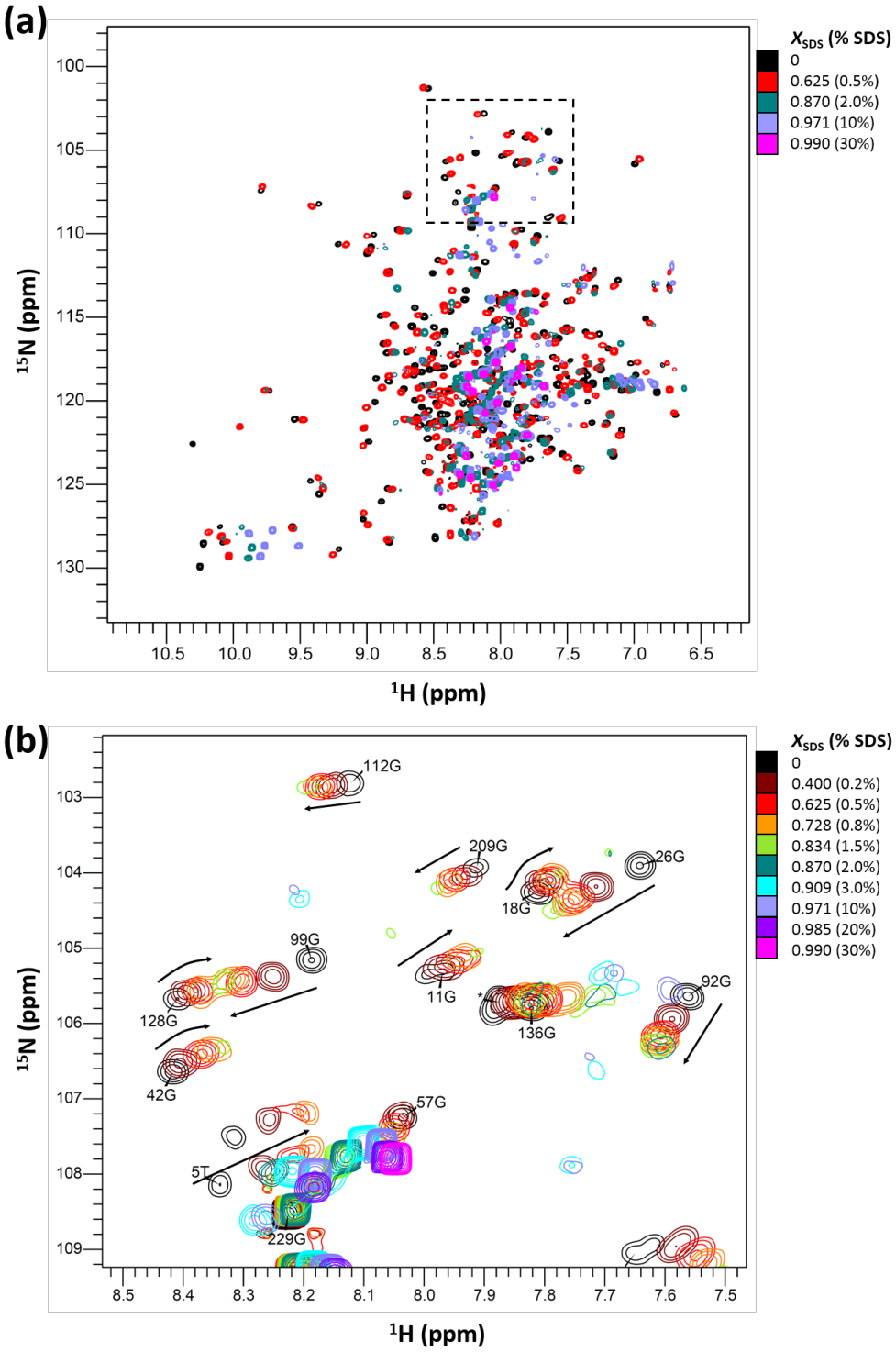


Figure 3.14: 2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC spectra of ~ 90 µM ^{15}N-labelled pSRII in 0–0.990 X_{SDS} (0–30% SDS in 0.5% c7-DHPC).
All spectra were recorded on an 800 MHz spectrometer at 308 K, with 240 scans. All samples contained TSP as internal reference. Panel **(b)** shows an expanded view of the glycine residues (boxed region of panel a). Chemical shift changes are represented using arrows next to each peak.

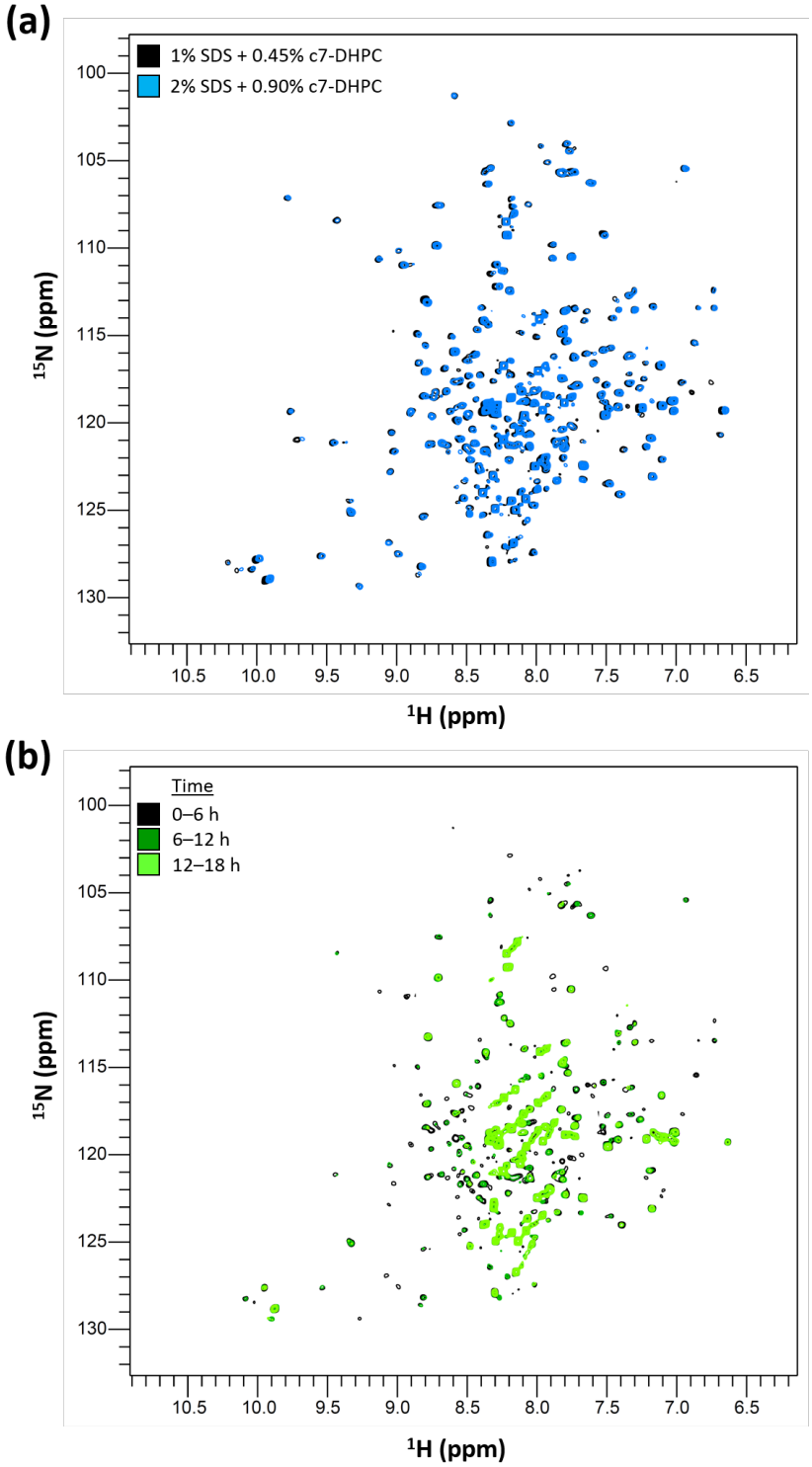


Figure 3.15: 2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC spectra recorded in different molar concentrations of SDS and c7-DHPC (a) and at different timepoints (b) in 0.834 X_{SDS} (1.5% SDS in 0.5% c7-DHPC)

All spectra were recorded on an 800 MHz spectrometer at 308 K, with 240 scans. All samples contained TSP as internal reference.

For a systematic understanding of structural changes of pSRII upon exposure to SDS, the combined chemical shift difference ($\Delta\delta(^{1}H,^{15}N)$) relative to pSRII in 0 X_{SDS} was calculated for each observable backbone amide signal across different X_{SDS} (Figure 3.16). $\Delta\delta(^{1}H,^{15}N)$ increases with increasing X_{SDS}, albeit to different extents for different residues. To identify individual residues or stretches of residues which experience substantial structural changes, residues with a large $\Delta\delta(^{1}H,^{15}N)$ are typically defined using a cut-off at one standard deviation above the average $\Delta\delta(^{1}H,^{15}N)$ across all residues [323]. Taking this cut-off, however, was not suitable for SDS-unfolded pSRII, as significant changes were only detected at the ends of helices (results not shown). On the one hand, this confirms the results from circular dichroism experiments, where the small changes in helicity were interpreted as fraying of helix ends and the transmembrane helices were thought to remain embedded in mixed micelles. On the other hand, this cut-off precludes a detailed study of the fine changes in tertiary structure occurring in the transmembrane region, including loosening of the retinal binding pocket and Schiff base hydrolysis. These considerations highlight the difficulties with interpreting backbone amide chemical shift changes, especially those from titration experiments using a molecule which interacts with the protein in a non-specific manner.

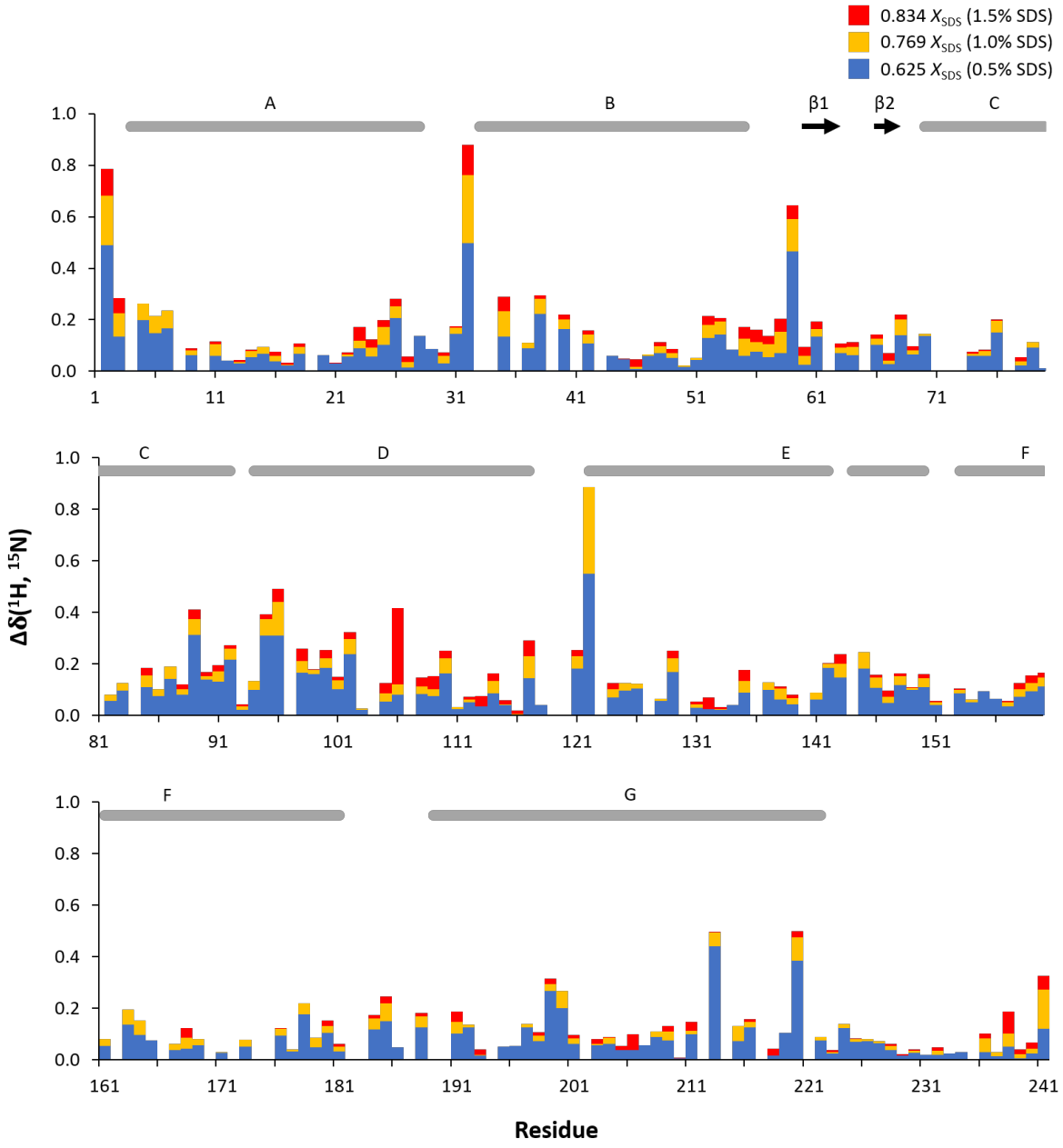


Figure 3.16: Combined chemical shift differences (Δδ(^{1}H,^{15}N)) of pSRII in 0.625, 0.769 and 0.834 X_{SDS}

Figure 3.17a shows Δδ(^{1}H,^{15}N) mapped onto the solution-state NMR structure of pSRII, with the values indicated by a gradual colour change from blue to red with every 0.1 ppm increment. In addition to identifying the big Δδ(^{1}H,^{15}N) occurring at helix ends, this approach has also enabled detailed study of fine structural changes in the transmembrane regions and the retinal binding pocket.

A 3 Å cut-off was taken around the retinal chromophore to identify residues which constitute the retinal binding pocket (Figure 3.17b). In 0–0.625 X_{SDS}, $\Delta\delta(^1H,^{15}N)$ of residues in the binding pocket are generally less than 0.1 ppm, indicating that there is little perturbation of the binding pocket. A notable exception is Trp-178, which has $\Delta\delta(^1H,^{15}N)$ between 0.1 and 0.2 ppm starting from 0.400 X_{SDS} (0.2% SDS in 0.5% c7-DHPC). This observation can be attributed to the residue being located near the extracellular end of helix F and is indicative of helix fraying rather than substantial disruption of the binding pocket. In ≥ 0.625 X_{SDS} (0.5% SDS in 0.5% c7-DHPC), several other residues show $\Delta\delta(^1H,^{15}N) > 0.1$ ppm. Starting from Trp-76 in the middle of helix C, residues with $\Delta\delta(^1H,^{15}N) > 0.1$ ppm gradually spread across helices C and D as X_{SDS} was increased, suggesting that the packing of helices C and D against the retinal binding pocket is beginning to loosen even before the unfolding transition (determined to be ~ 0.81 X_{SDS} for SDS + 0.5% c7-DHPC in Chapter 2). $\Delta\delta(^1H,^{15}N) > 0.1$ ppm was detected for Phe-208 and Gly-209 in helix G starting from 0.728 X_{SDS} and 0.769 X_{SDS}, respectively. Both residues are located around one helical turn away from the site of retinal attachment (Lys-205), indicating that small structural changes were beginning to appear near the retinal Schiff base before the unfolding transition. In comparison, $\Delta\delta(^1H,^{15}N)$ remained generally small (< 0.1 ppm) for other residues in helices F and G (including Lys-205) within the retinal binding pocket. In 0.834 X_{SDS} (just beyond the unfolding transition), Val-206 experienced $\Delta\delta(^1H,^{15}N) > 0.1$ ppm whereas Lys-205 still retained $\Delta\delta(^1H,^{15}N) < 0.1$ ppm. Hence, although retinal Schiff base hydrolysis can be associated with small structural changes in helix G up to one helical turn away from the Schiff base, Lys-205 itself experienced little structural changes. It is noted that Lys-205 was unobservable beyond 0.834 X_{SDS}. In fact, the number of observable residues gradually decreased from 0.728 X_{SDS} onwards, showing the necessity of studying changes in signal intensities and protein backbone dynamics (Sections 3.3.3–3.3.5).

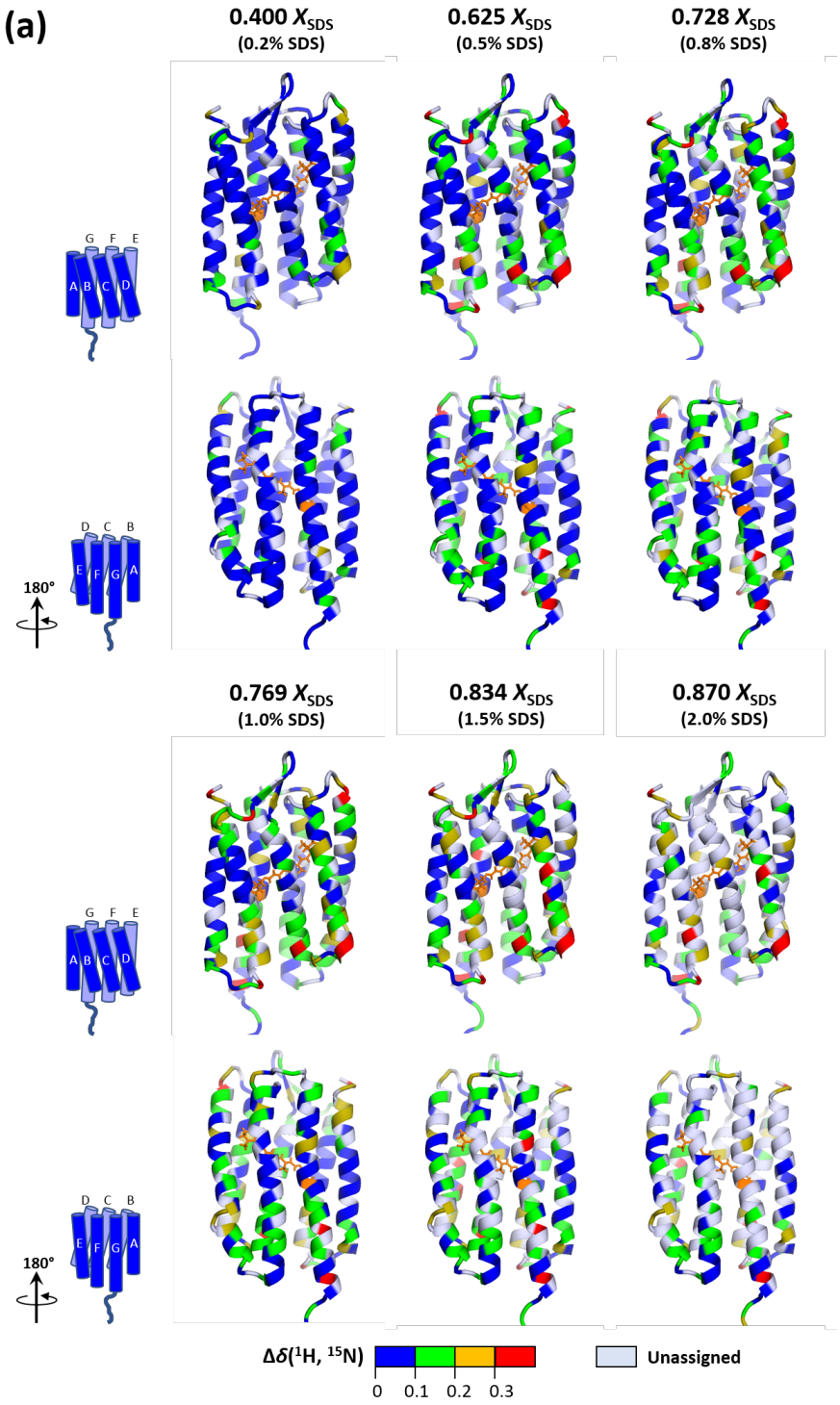


Figure 3.17: Δδ(^{1}H,^{15}N) of pSRII backbone amides in different X_{SDS}
Δδ(^{1}H,^{15}N) are mapped onto the solution-state NMR structure of pSRII (PDB 2KSY) [216]. **(a)** Residues are coloured from blue to red according to Δδ(^{1}H,^{15}N) values, showing greater structural changes at the ends of helices than in the transmembrane region. Unassigned residues are coloured white.

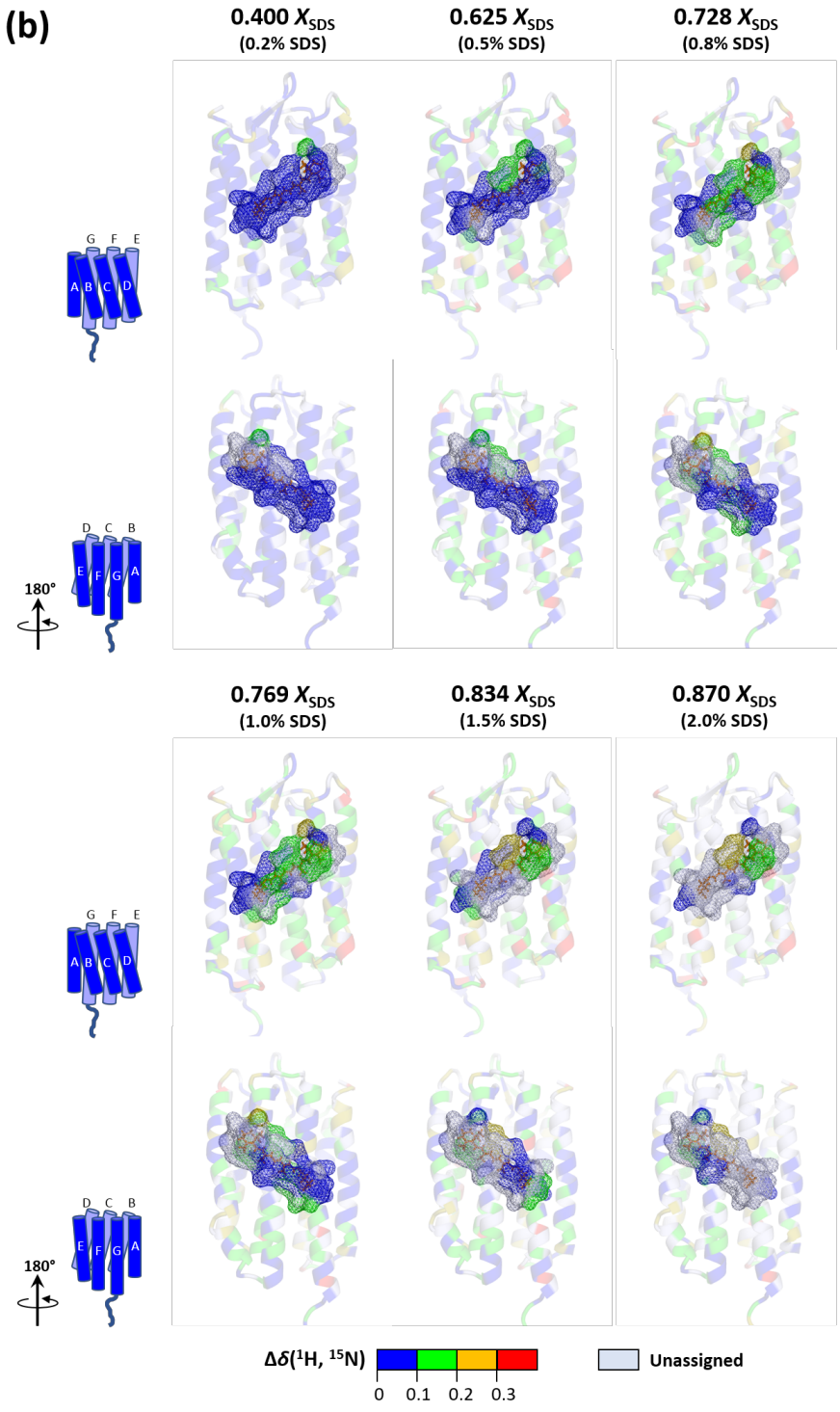


Figure 3.17: Δδ(^{1}H,^{15}N) plotted on the PDB structure of pSRII (cont.)
(b) Surface representation of residues within 3 Å of retinal coloured according to Δδ(^{1}H,^{15}N) values to show structural changes at the retinal binding pocket.

Overall, analysis of backbone amide chemical shift changes suggests that disruption of the binding pocket initiates with disrupted packing of helices C and D against retinal before the unfolding transition followed by small changes in helix G near the site of retinal attachment both before and beyond the unfolding transition. Conformational changes in other parts of the binding pocket were limited in comparison.

Under the fast exchange regime, the presence of more than two exchanging conformations in SDS-unfolded pSRII can be detected by examining the trajectories of peak shifts [323]. Many residues showed non-linear peak shift trajectories, indicating the presence of more than two species which are undergoing conformational exchange. Ser-44 is shown as an example in Figure 3.18a: a linear trajectory was observed in 0–0.625 X_{SDS}, indicating fast-to-intermediate conformational exchange between the folded and unfolded conformations and an increase in the population of the unfolded conformation with increasing X_{SDS}. The peak showed little to no changes in chemical shift between 0.625 and 0.769 X_{SDS}, suggesting that the chemical environment of Ser-44 remains similar in this X_{SDS} range. Ser-44 was unobservable in 0.834–0.870 X_{SDS}, indicating exchange broadening. The peak reappeared in 0.893 and 0.909 X_{SDS}, with peak shifts showing a different trajectory to that observed in 0–0.625 X_{SDS}, indicating that further increase in X_{SDS} led to the formation of alternative conformation(s) and its exchange with the conformation observed in 0.625–0.769 X_{SDS}. Such complex peak trajectories were only observed for a very limited number of residues, and no further conclusions could be drawn. Backbone amides with non-linear peak trajectories are mapped onto the solution-state NMR structure of pSRII in Figure 3.18b. Residues showing conformational exchange between more than two conformations are spread out across all seven transmembrane helices. This suggests that SDS-mediated unfolding of pSRII has led to widespread “softening” of the transmembrane helical bundle, as the ends of the helices are beginning to unravel (helix fraying) and the tertiary structure becomes less tightly packed.

In summary, pSRII progressively unfolds in increasing X_{SDS} to yield an ensemble of different unfolded conformations.

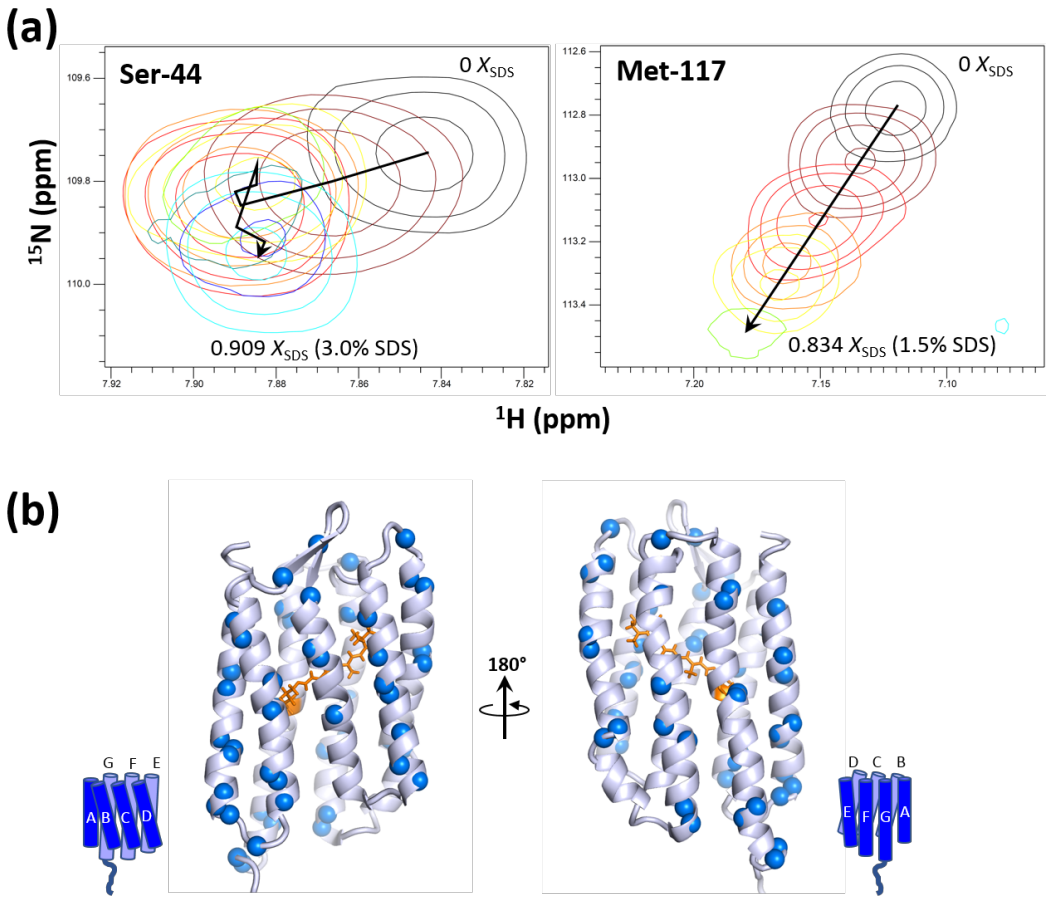


Figure 3.18: Non-linear peak shift trajectories indicate exchange amongst more than two conformations
(a) Examples of peaks exhibiting non-linear peak shift trajectories (Ser-44) vs. a linear trajectory (Met-117). **(b)** Backbone amides with non-linear peak shift trajectories are mapped onto the pSRII structure as blue spheres, showing widespread conformational exchange across the protein.

3.3.3 Peak intensities: Backbone dynamics of SDS-denatured pSRII

Decreases in peak intensities with increasing X_{SDS} led to difficulties with analysing chemical shift changes due to fewer observable signals. 1D ^{1}H NMR spectra already illustrate that pSRII in ≤ 0.769 X_{SDS} showed no changes in peak intensities over time, whereas intensity loss over time was observed in spectra of pSRII beyond the unfolding transition (in ≥ 0.834 X_{SDS}) (Figure 3.11). These observations indicate that backbone amide conformational exchange in the µs–ms timescale only occur in unfolded pSRII. Detailed understanding of peak intensity changes would yield information on how changes in backbone dynamics and conformational exchange relate to pSRII unfolding.

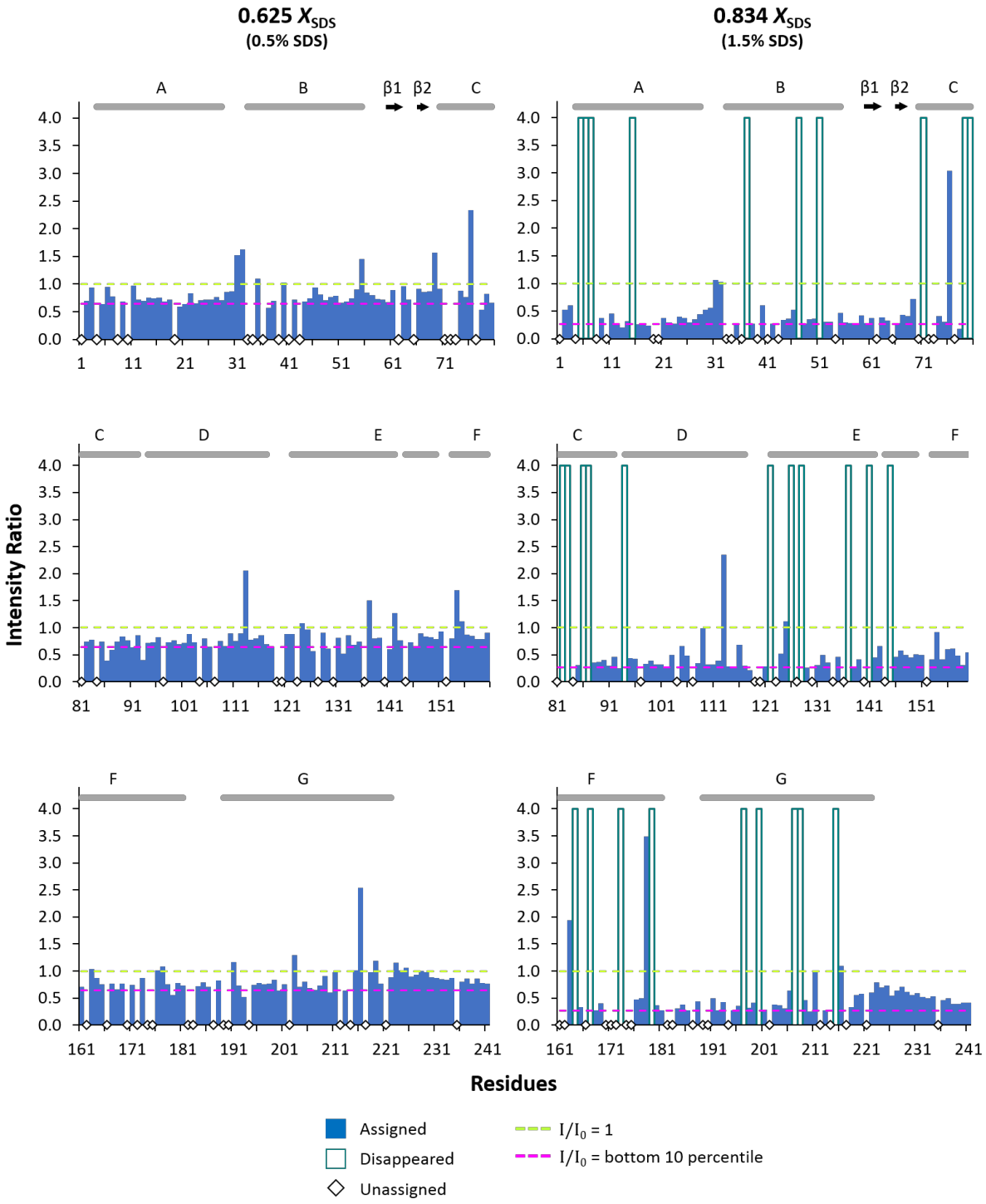


Figure 3.19: Intensity changes in 0.625 and 0.834 X_{SDS}
Intensity changes are shown as I/I_0, the ratio between peak intensity at a given X_{SDS} relative to the intensity at 0 X_{SDS}, and are represented as blue bars. Unobservable residues are represented as unfilled bars of arbitrary height. Unassignable residues are shown as unfilled diamonds. The cut-off values of $I/I_0 = 1$ and I/I_0 at the bottom 10^{th} percentile are shown as lime and magenta dashed lines, respectively. α-helices and β-sheet secondary structures are shown as grey rods and black arrows, respectively, at the top of each plot.

Changes in peak intensities were calculated by taking the ratio of the peak intensity at a given X_{SDS} against the peak intensity in 0 X_{SDS}:

$$I/I_0 \text{ (intensity ratio)} = \frac{\text{Peak intensity at } n\ X_{SDS}}{\text{Peak intensity at } 0\ X_{SDS}} \quad (3.48)$$

Peak intensities in the presence of SDS were generally lower than those in 0 X_{SDS} ($I/I_0 < 1$), and decreased with increasing X_{SDS} (Figure 3.19). To identify peaks with big changes in intensities, cut-off values at $I/I_0 > 1$ or I/I_0 below the 10th percentile were applied. A few individual residues in the middle of helices C, D, E and G and in the loop regions showed $I/I_0 > 1$, indicating that these backbone amides became more flexible in the presence of SDS. In 0.625 X_{SDS}, residues with I/I_0 below the 10th percentile were mostly located in helix C and the extracellular half of helix E. This appears to be consistent with the relatively greater number of unobservable residues in these two helices compared to other helices at 0.834 X_{SDS} (seven residues in helix C and six residues in helix E). Figure 3.19 did not show any obvious stretches of residues with prominent intensity changes, possibly suggesting that significant changes in backbone dynamics might be occurring at only several specific residues which are spatially (rather than sequentially) close to each other.

To better visualise the distribution of residues with prominent intensity changes, I/I_0 are mapped onto pSRII and coloured by value (Figure 3.20). Residues with $I/I_0 < 1$ are further classified as being above or below the average taken across all helical residues. Solvent-exposed residues, determined by titration with the spin label reagent gadoteridol [216], are highlighted in blue in Figure 3.20a. Most of the micelle-embedded residues showed below-average I/I_0, indicating that backbone amides in the transmembrane region are less flexible than those in the loop regions. A few residues at the micelle/solvent interface showed above-average I/I_0. Residues with $I/I_0 > 1$ appear to form clusters. In 0.625 X_{SDS}, most residues with $I/I_0 > 1$ are solvent-exposed or located at the micelle/solvent interface. There are two notable exceptions: Trp-76 in the middle of helix C and Val-203 in the middle of helix G, both of which are located within the binding pocket. It must be emphasised that the reported peak intensities for Trp-76 contained a small contribution from Thr-218 in $\geq$ 0.728 X_{SDS} due to overlap of the two peaks in the spectra. Nevertheless, it is clear from assessing the peak shapes that Trp-76 remained observable up to 0.870 X_{SDS}. A few residues

within the binding pocket on helices D and G also show above-average I/I_0, possibly suggesting preliminary loosening of the binding pocket. This is further supported by residues with above-average I/I_0 being found in the same regions of helices D and G in 0.834 X_{SDS}.

In 0.834 X_{SDS}, there was a marked decrease in the number of residues with $I/I_0 > 1$, indicating that there is a general decrease in peak intensities across the protein, including residues in solvent-exposed regions and micelle/solvent interfaces. Interestingly, most residues in helix D with above-average I/I_0 (Asn-105 & Met-109) or $I/I_0 > 1$ (Phe-113) are facing the interface between helices C and D. Asn-105 and Met-109 showed further increase in peak intensities to yield $I/I_0 > 1$ in 0.870 X_{SDS}, indicating increase in backbone amide flexibility and gradual disruption of inter-helical interactions between helices C and D. In contrast, most residues on helix C were unobservable in 0.870 X_{SDS}. This indicates that backbone amides on adjacent faces of the two helices show different degrees of flexibility, hence lending further support towards the proposal of inter-helical interactions between helices C and D being disrupted.

Residues within the binding pocket (Figure 3.20b) also showed a general decrease in peak intensities, and the number of observable residues also decreased beyond the unfolding transition. Between 0.834 and 0.870 X_{SDS}, residues disappeared from the cytoplasmic halves of helices B and C, the middle of helix E, and the extracellular half of helix F. On the other hand, residues on helix G packing against Lys-205, the Schiff base and up to C13–C14 of the retinal polyene chain remained observable at 0.834 X_{SDS}, albeit with low peak intensities (below average). These observations indicate that at just above the unfolding transition, conformational exchange occurs throughout the binding pocket but to a smaller extent for residues surrounding the Schiff base, suggesting that the Schiff base is disrupted later compared to the rest of the binding pocket. In 0.870 X_{SDS}, observable residues in the binding pocket were mostly clustered on helix D and showed $I/I_0 > 1$, further indicating that the extracellular half of helix D exhibits faster backbone dynamics in high X_{SDS}.

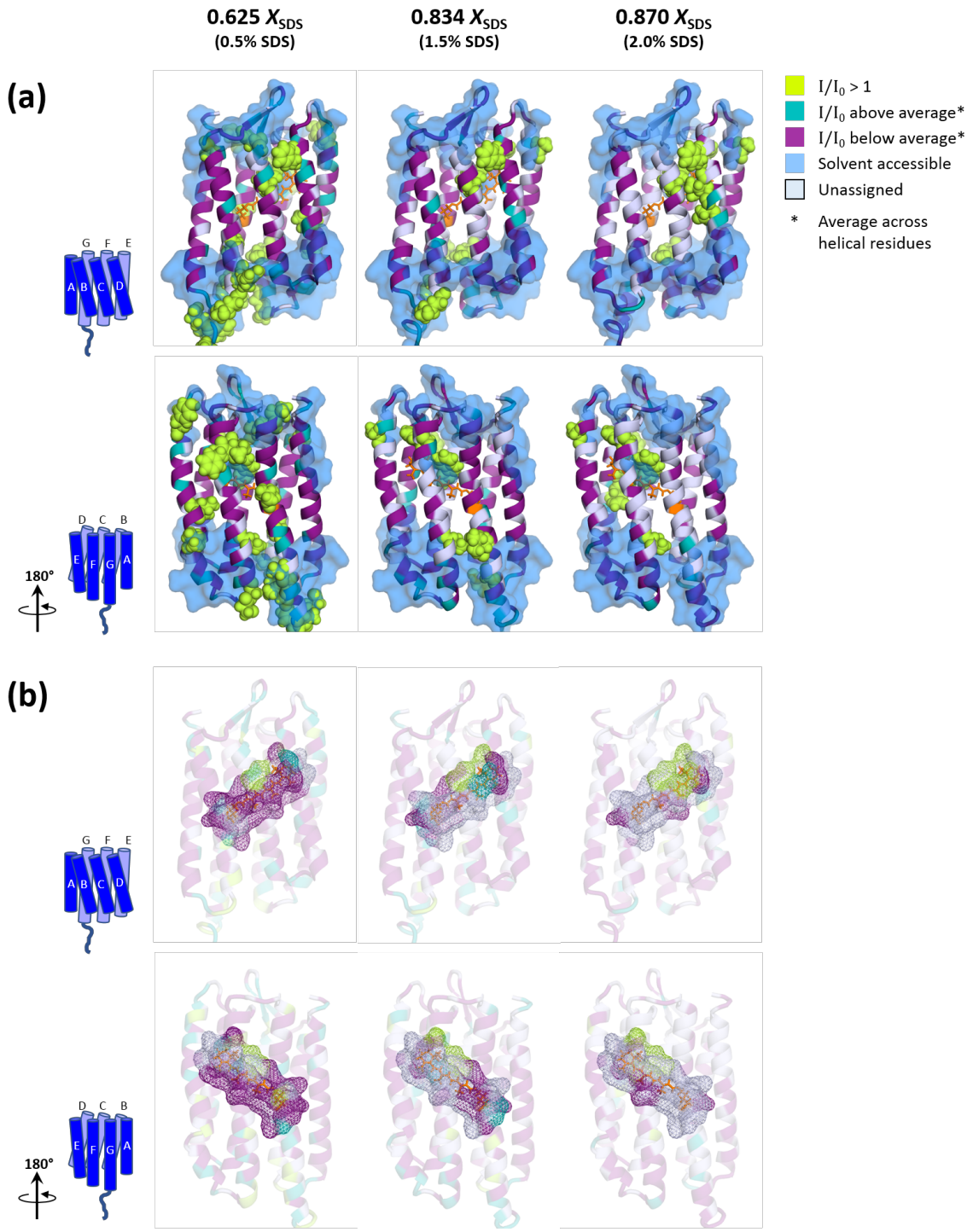


Figure 3.20: Intensity ratios mapped onto pSRII structure
(a) Residues with $I/I_0 > 1$ are shown as lime spheres. Other residues are coloured according to I/I_0 values. An average I/I_0 value is calculated across helical residues. Residues with below-average I/I_0 are coloured in purple, and residues with above-average I/I_0 in teal. Solvent-exposed regions, determined from NMR signal attenuation in the presence of the soluble spin label reagent gadoteridol [216], are shaded in pale blue. **(b)** Surface representation of residues within 3 Å of retinal coloured according to I/I_0 values to show changes in pSRII backbone dynamics at the retinal binding pocket.

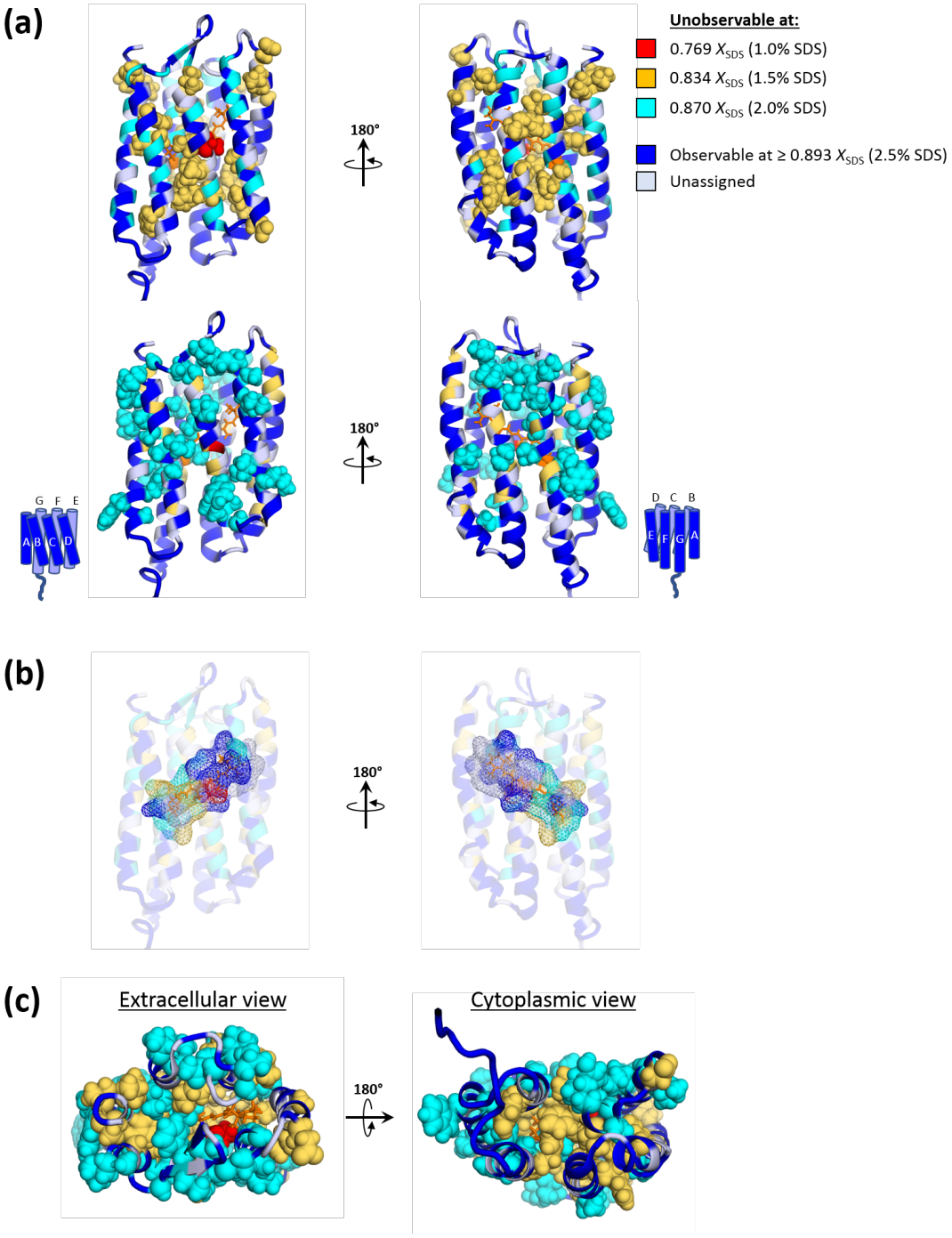


Figure 3.21: Unobservable residues at different X_{SDS} mapped onto pSRII structure
(a) Residues which disappear at 0.769, 0.834 and 0.870 X_{SDS} are shown as red, yellow or cyan spheres, respectively. Residues which remain observable at ≥ 0.893 X_{SDS} are coloured in blue. **(b)** Surface representation of residues within 3 Å of retinal coloured according to X_{SDS} at which the residues become unobservable, showing changes in pSRII backbone dynamics at the retinal binding pocket. **(c)** Extracellular and cytoplasmic views, showing a greater proportion of inward-facing helical residues amongst residues which disappear at lower 0.769 and 0.834 X_{SDS}. Residues are represented using the same scheme as in panel a.

Disappearing peaks are interpreted as intermediate conformational exchange in the μs–ms timescale leading to line broadening and very low peak intensities. The first unambiguous peak disappearance (Thr-80) occurred at 0.728 X_{SDS}. Unassignable residues in 0.400 and 0.625 X_{SDS} were assumed to either be significantly overlapped with other peaks or unassignable due to intractably large chemical shift changes. Residues which became unobservable in 0.769–0.870 X_{SDS} are mapped onto the pSRII structure and shown as coloured spheres (Figure 3.21).

The number of unobservable residues increased with increasing X_{SDS}. Thr-80 (red spheres) on helix C, which packs against C9–C12 in the middle of the retinal polyene chain, was the first residue to become unobservable at 0.728 X_{SDS}. In 0.834 X_{SDS} (just above the unfolding transition), a major inter-helical cluster was formed between unobservable residues in the cytoplasmic halves of helices C, E, F and G, and much smaller clusters were formed between unobservable residues in the extracellular halves of helices A and B and between helices F and G. The chi-squared test (Table 3.1) shows that residues which have become unobservable at 0.769 X_{SDS} and 0.834 X_{SDS} are significantly populated in helical regions (P = 0.026). Out of such helical residues, there is significant preference for residues which face inwards towards the protein core (P = 0.003). This indicates that the backbone amides of helical residues facing the protein core were the earliest to experience prominent conformational exchange in the μs–ms timescale. Within the binding pocket, unobservable residues include Met-15 from helix A, Ala-47 from helix B, Thr-79 and Ile-83 from helix C and Gly-207 and Phe-208 from helix G, all of which are located near the Schiff base, packing against Lys-205 and up to C13–C14 of the retinal polyene chain. Tyr-51 from helix B, which is involved in the hydrogen-bond network of ordered water molecules in pSRII (PDB 1H68; [336]), also disappeared in 0.834 X_{SDS}. To summarise, in 0.834 X_{SDS} (just above the unfolding transition), prominent amounts of conformational exchange in the μs–ms timescale was observed for residues involved in inter-helical interactions amongst the cytoplasmic halves of helices C, E, F and G, residues around the Schiff base, and a small number of residues on the extracellular side of pSRII. It could therefore be hypothesised that loosening of inter-helical packing is initiated from the cytoplasmic side, likely amongst helices C, E, F and G, and propagated towards the Schiff base.

Residues which became unobservable at 0.870 X_{SDS} were spread across the entire protein. The slight bias in the distribution of such residues towards the extracellular side could be because many residues in the cytoplasmic side have already disappeared in lower X_{SDS}. The chi-squared test showed no significant difference in the distribution of unobservable residues between helical vs. loop residues or helical residues which face inwards (towards protein core) or outwards (towards detergent micelle). Except for Gly-112 on helix D, residues within the binding pocket which became unobservable at 0.870 X_{SDS} were solely found on helix G, including Asp-201, Thr-204, Lys-205, Val-206 and Gly-209, all of which are within one helical turn away from the Schiff base. Asp-201 and Thr-204 are also involved in the water network, possibly suggesting that the water network is also being progressively disrupted with increasing X_{SDS}.

	0.769 & 0.834 X_{SDS}	**0.870 X_{SDS}**
Helical vs. Loop	P = 0.026 (*)	P = 0.093
Inward-facing vs. Outward facing (helical residues only)	P = 0.003 (*)	P = 0.481

Table 3.1: *P*-values from chi-squared test on the distribution of residues which have become unobservable at different X_{SDS}

Combining the observations from analysing chemical shift changes, peak intensity ratios and peak disappearance, it is interpreted that loosening of the binding pocket before the unfolding transition involves widespread backbone amide conformational exchange in the μs–ms timescale throughout the binding pocket. Residues in helices C and D showed disrupted packing against retinal (chemical shift changes). Residues in helix D showed increased backbone flexibility ($I/I_0 > 1$). Inter-helical interactions between helices C and D were likely disrupted (different backbone dynamics). Prominent conformational exchange (peak disappearance) of backbone amides first began in the cytoplasmic halves of helices C, E, F and G. Disruption of the binding pocket was along the retinal polyene chain with the disappearance of Thr-80 packing against the middle (C9–C12) of retinal followed by several residues from helices A, B, C and G which pack against C13–C14 of retinal and Lys-205. The middle of helix G in general, including Lys-205 and neighbouring residues, showed smaller amounts of μs–ms timescale conformational exchange in 0.834 X_{SDS} compared to the

rest of the binding pocket. Except for the aforementioned residues in helices C and D, these residues on helix G near Lys-205 only became unobservable at 0.870 X_{SDS}. These results indicate that the Schiff base and neighbouring residues on helix G were the last parts of the binding pocket to be disrupted. Schiff base hydrolysis, which occurs from 0.834 X_{SDS} onwards, was also accompanied by small structural changes in helix G up to one helix turn away from the Schiff base. Lys-205 itself experienced little changes in chemical environment, and hence changes in backbone flexibility likely had a greater contribution towards Schiff base hydrolysis. While the analysis of backbone amide peak disappearance points towards progressive disruption of the hydrogen-bond network of ordered water molecules in pSRII, it must be noted that since the hydrogen-bond network is held together by side-chains rather than backbone amides, changes detected in backbone amides might be small and could contain other contributions.

While most backbone amides experience predominantly conformational exchange in the intermediate (μs–ms) timescale leading to line broadening, the backbone amides of at least two residues exhibit conformational exchange at more than one timescale (Figure 3.22a). In 0.728–0.834 X_{SDS}, the backbone amides of Val-17 and Phe-98 each had two resonances, indicating slow exchange between the two corresponding backbone conformations. A shift in the relative populations of the two conformations can be seen between 0.728 and 0.834 X_{SDS}. In $\geq$ 0.870 X_{SDS}, both peaks of the two residues showed decreased intensities, indicating a transition from slow exchange towards intermediate exchange as the peaks became broadened beyond detection. From 0.909 X_{SDS} onwards, several other backbone amides were also found to undergo slow exchange, as evidenced by difficulties in assigning several peaks in the glycine region of the spectrum due to big chemical shift changes (Figure 3.22b). The conclusions which could be drawn from Val-17 and Phe-98 were limited by the small number of residues involved and their lack of obvious functional relevance (i.e. not in the binding pocket). Nevertheless, these observations highlight the heterogeneous nature of the ensemble of unfolded states with different modes of backbone amide conformational exchange being observed at different X_{SDS}.

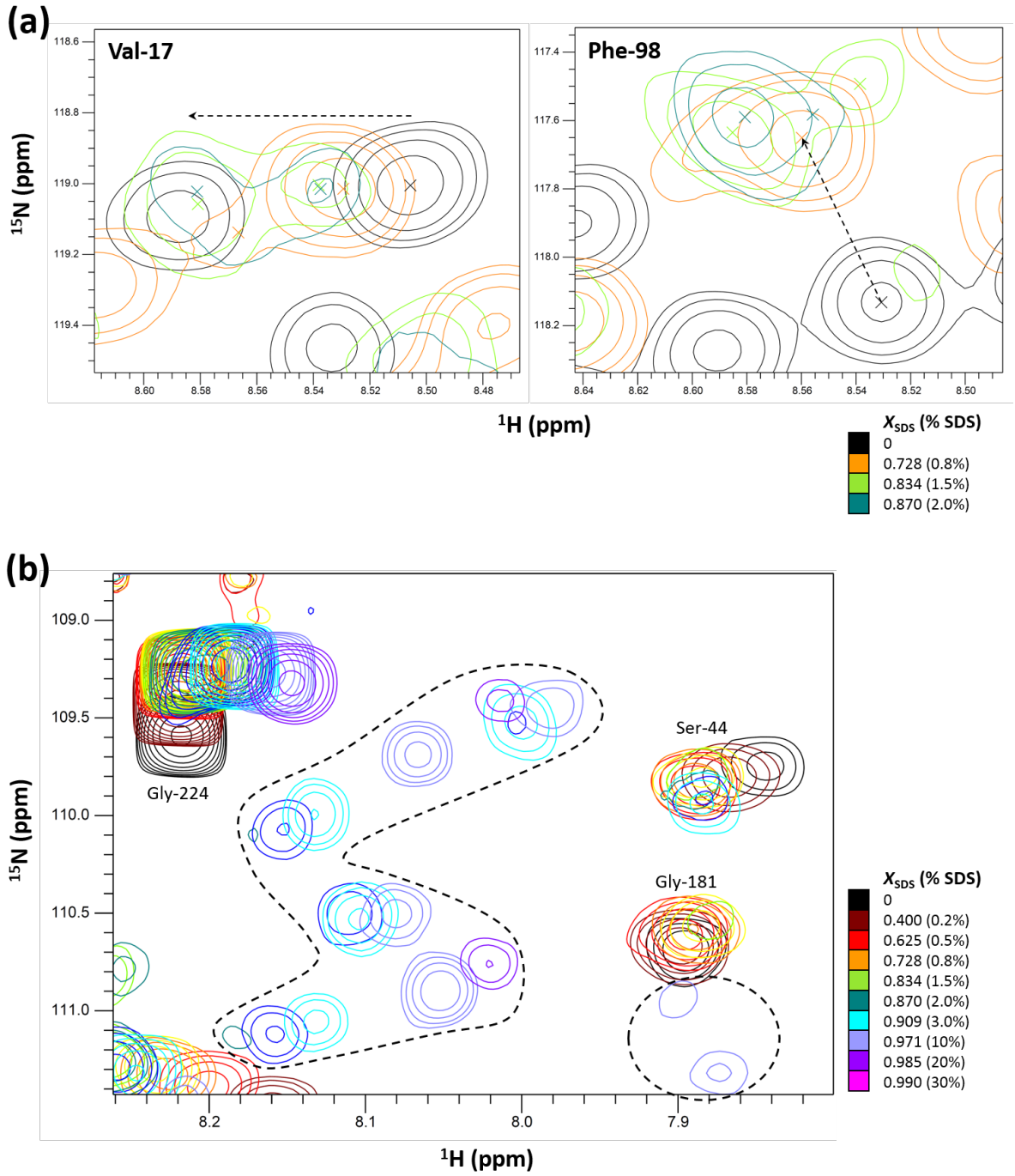


Figure 3.22: Examples of residues showing changes in conformational exchange regime with increasing X_{SDS}

(a) Val-17 and Phe-98 backbone amides (peaks with crosses) show transition from fast exchange regime in 0 X_{SDS} to slow exchange regime in 0.728–0.870 X_{SDS} to intermediate exchange regime in > 0.870 X_{SDS}. Chemical shift changes are represented using dotted arrows. **(b)** Several backbone amides (in regions defined by dotted lines) show slow timescale conformational exchange and big chemical shift differences in high X_{SDS}, leading to difficulties with peak assignment.

3.3.4 Solvent accessibility: Hydrogen/deuterium exchange

Tryptophan fluorescence studies in Chapter 2 showed that changes in tertiary structure also involve increases in solvent accessibility of certain tryptophan residues. Hydrogen/deuterium (H/D) exchange measurement by NMR was pursued to explore changes in solvent accessibility of backbone amides across pSRII. Spectra were measured in protonated solvent (9% D_2O was included for lock stabilisation) or deuterated solvent (90% D_2O). Concentrated SDS stock (25% SDS + 0.5% c7-DHPC in protonated or deuterated solvent) was added directly to pSRII in 0 X_{SDS} to yield 0.769 X_{SDS} (1% SDS + 0.5% c7-DHPC). Measurements must be done in low X_{SDS} (below the unfolding transition) to ensure that loss in peak intensity is solely due to deuterium incorporation, without contributions from changes in backbone dynamics or aggregation over time (experimental time = 6 h).

Deuterium incorporation was quantified for each assignable backbone amide in 0 X_{SDS} and 0.769 X_{SDS} by calculating the intensity ratio in deuterated solvent vs. in protonated solvent (I_D/I_H). Comparison of I_D/I_H in 0 X_{SDS} (white bars) and 0.769 X_{SDS} (orange dots) plotted in Figure 3.23a shows that peak intensity loss was smaller in 0.769 X_{SDS}, meaning less deuterium has been incorporated and solvent accessibility is generally lower in the presence of SDS. This difference in deuterium incorporation is more prominent in transmembrane helical regions, possibly because of a "shielding" effect conferred by the presence of more detergent molecules and larger micelle sizes. Several residues are noted to have greater deuterium incorporation, and hence greater solvent accessibility, in 0.769 X_{SDS}. For easier visualisation of different degrees of deuterium incorporation, the difference in intensity ratios, $\Delta(I_D/I_H)$, was calculated as shown in Equation 3.49 and plotted in Figure 3.23b.

$$\Delta\left(\frac{I_D}{I_H}\right) = \left(\frac{I_D}{I_H}\right)_{0.769} - \left(\frac{I_D}{I_H}\right)_0 \quad (3.49)$$

Positive values indicate greater solvent accessibility in 0 X_{SDS}; negative values indicate greater solvent accessibility in 0.769 X_{SDS}. The average $\Delta(I_D/I_H)$ across helical residues was 0.045, meaning on average helical residues showed less solvent accessibility in 0.769 X_{SDS}.

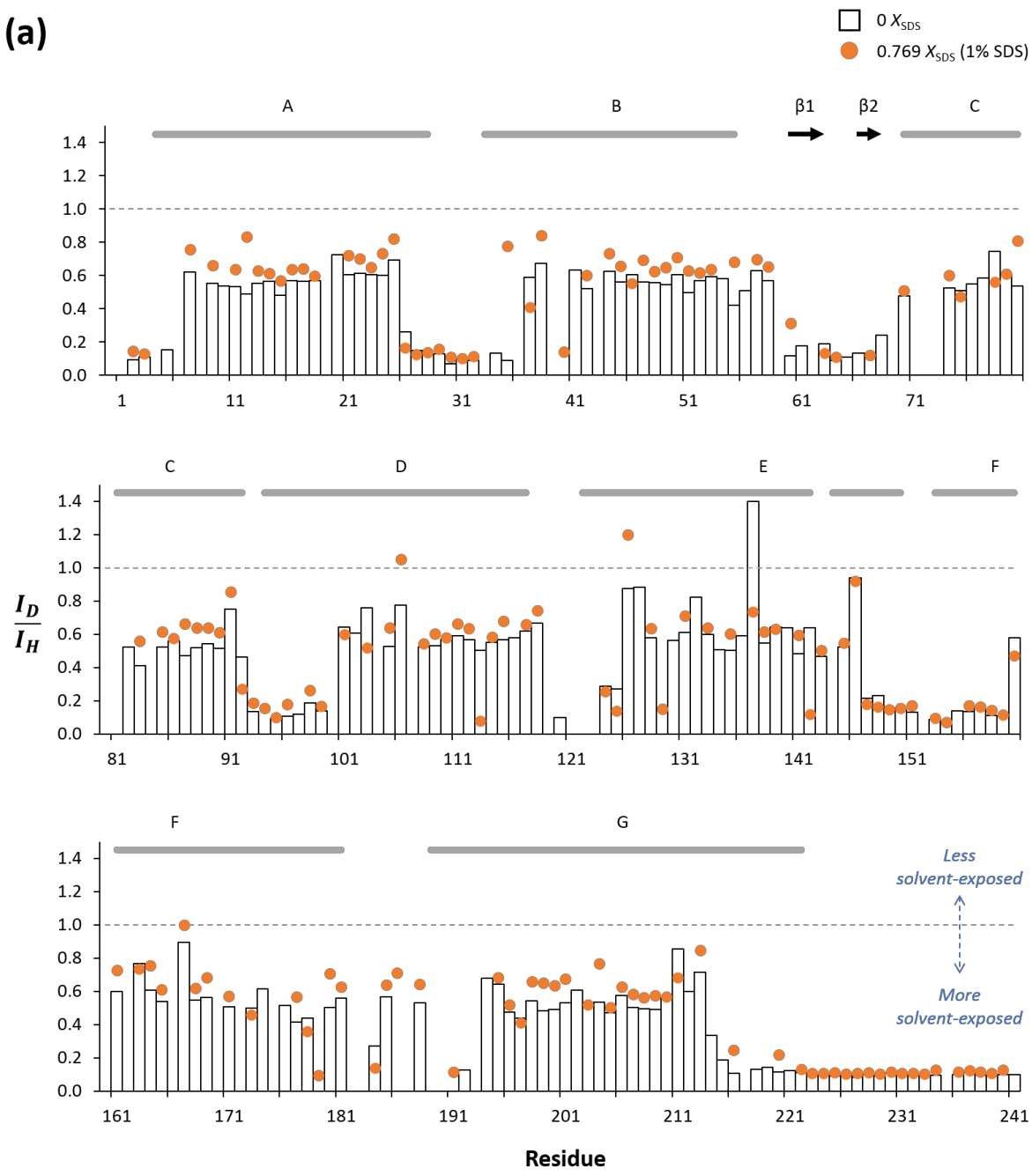


Figure 3.23: Hydrogen/deuterium exchange in 0 X_{SDS} vs. 0.769 X_{SDS}
(a) Deuterium incorporation is quantified by the ratio of peak intensities measured in deuterated solvent vs. in protonated solvent (I_D/I_H), where higher I_D/I_H indicates less deuterium incorporation and hence less solvent exposure. Comparison of I_D/I_H in 0 X_{SDS} (white bars) vs. in 0.769 X_{SDS} (orange dots) shows that backbone amides are generally less solvent-exposed in the presence of SDS.

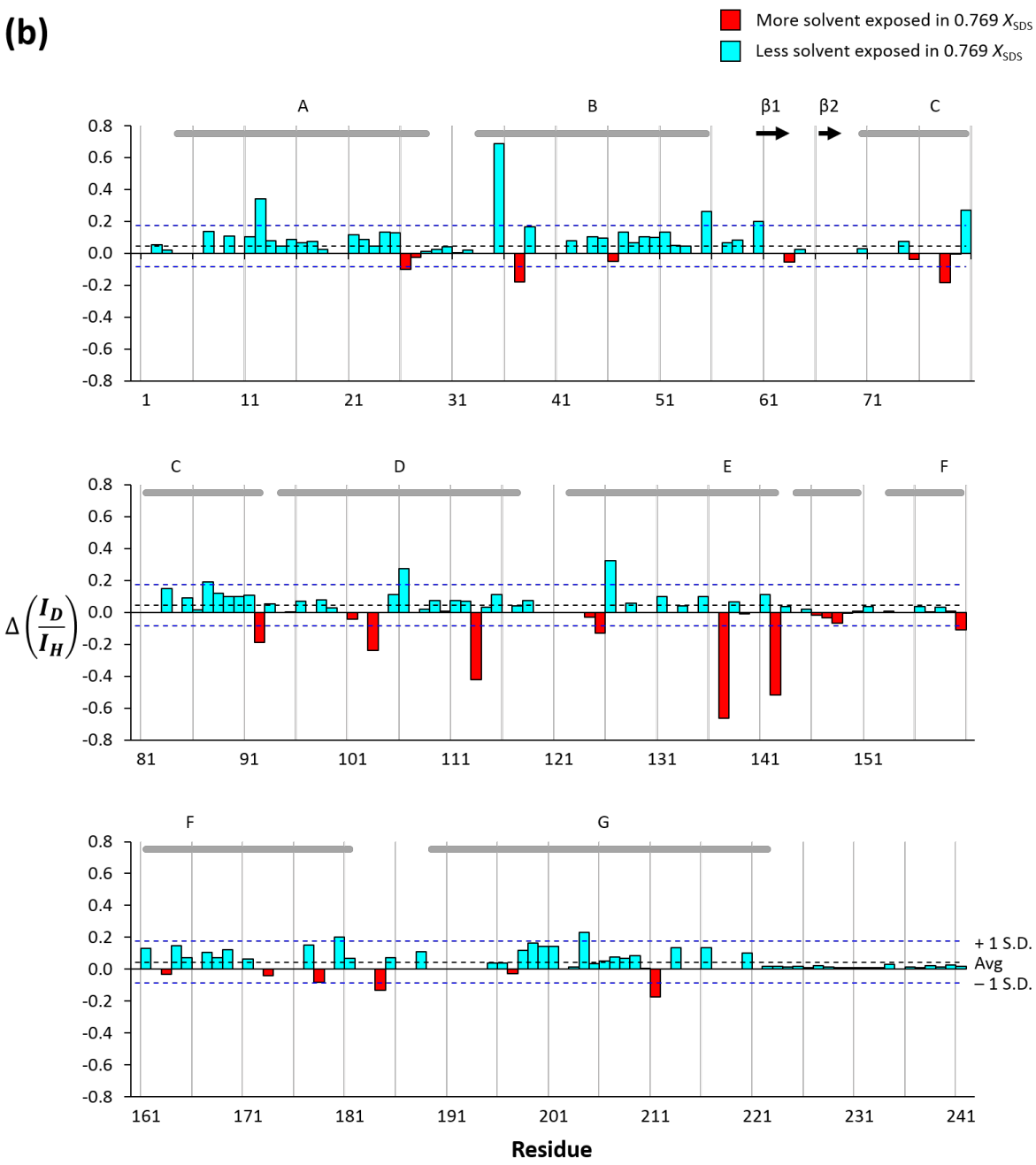


Figure 3.23: Hydrogen/deuterium exchange in 0 X_{SDS} vs. 0.769 X_{SDS} (cont.)
(b) The difference in I_D/I_H at different X_{SDS}, $\Delta(I_D/I_H)$, is calculated for each assigned residue and coloured by value. Residues which are more solvent-exposed in 0.769 X_{SDS} (negative $\Delta(I_D/I_H)$) are coloured in red; residues which are less solvent-exposed in 0.769 X_{SDS} (positive $\Delta(I_D/I_H)$) are coloured in cyan. Residues with significant changes in I_D/I_H are identified by taking cut-off values of one standard deviation above or below the mean $\Delta(I_D/I_H)$ taken across helical residues.

To identify residues with significant differences in solvent accessibility between different X_{SDS}, cut-off values were taken at one standard deviation above or below the average taken across the helical residues. Such residues are mapped onto the pSRII structure as coloured spheres, and the other residues are coloured according to $\Delta(I_D/I_H)$ value (Figure 3.24). Residues with good solvent accessibility (unobservable in deuterated solvent) in both 0 X_{SDS} and 0.769 X_{SDS} are represented as grey spheres; these residues are located exclusively in solvent exposed regions.

Most residues within the binding pocket showed less solvent accessibility in 0.769 X_{SDS}. Thr-80 in the middle of helix C and Thr-204 in the middle of helix G were the only two residues with significant differences in solvent accessibility (greater than one standard deviation away from the mean), both of which became less solvent accessible in the presence of SDS. Trp-178 on the extracellular end of helix F and Thr-79 in the middle of helix C experienced minor increases in solvent accessibility. Trp-178 is located at the micelle/solvent interface, hence explaining its slight increase in solvent exposure due to helix fraying. Thr-79 is hydrogen-bonded to Asp-75, which is the Schiff base counterion and hydrogen-bonded to the water network via its side-chain. Asp-75 also showed a small increase in solvent accessibility, suggesting that the water network might have started to loosen.

To summarise, the presence of SDS has led to a general decrease in solvent accessibility of backbone amides in the transmembrane region, likely reflecting the increase in detergent concentration. A few residues show increased solvent accessibility in 0.769 X_{SDS}, most of which are located at the micelle/solvent interface or at solvent-accessible regions. Slight increases in solvent accessibility observed for residues which are hydrogen-bonded to the internal water network could suggest preliminary loosening of the water network. Changes in solvent accessibility are small in the transmembrane region and within the binding pocket, consistent with the tertiary structure of pSRII remaining largely intact in 0.769 X_{SDS} (i.e. before the unfolding transition; see Chapter 2).

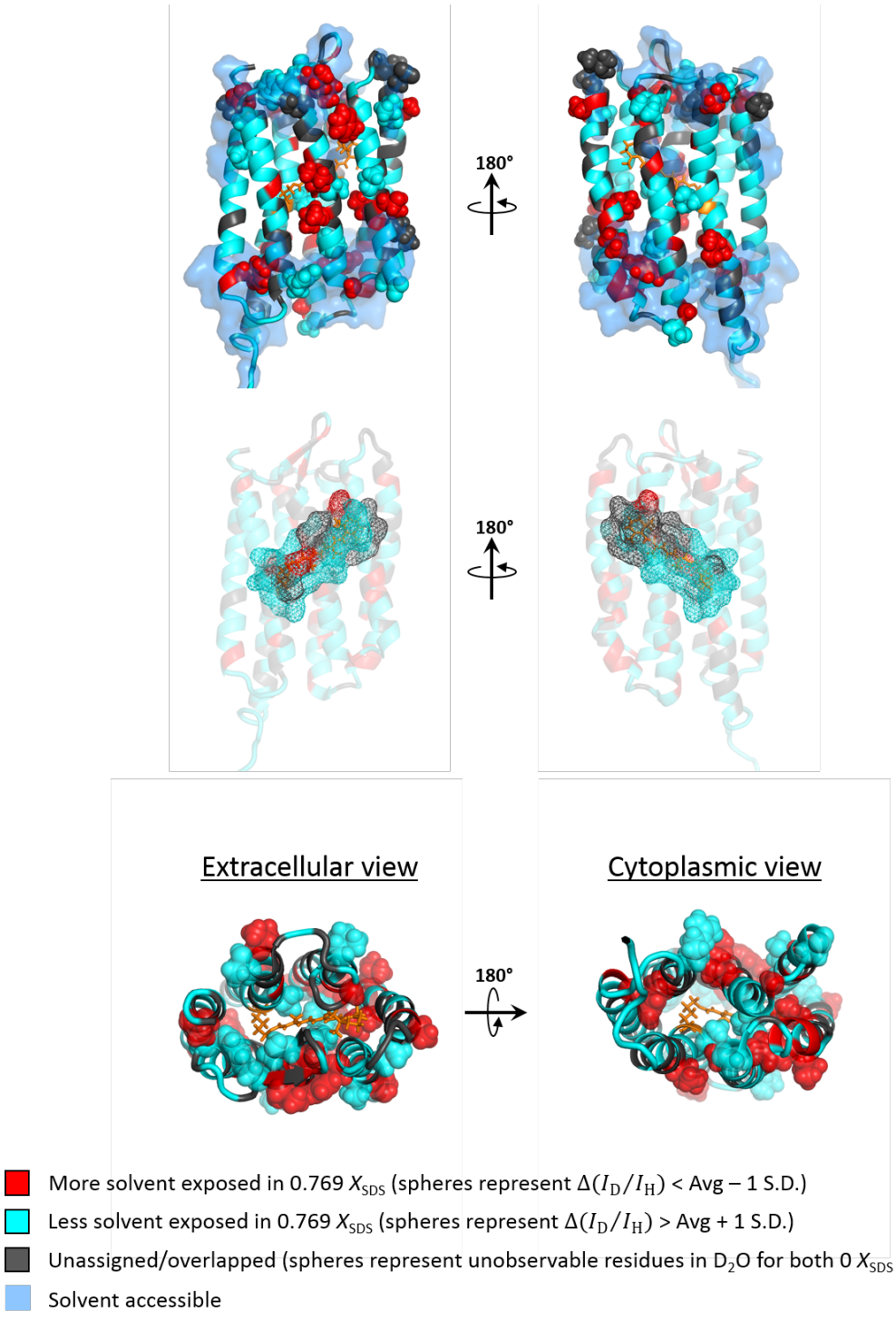


Figure 3.24: Differences in solvent accessibility in 0 X_{SDS} vs. 0.769 X_{SDS} mapped onto pSRII structure

Difference in solvent accessibility is quantified for each residue using the parameter $\Delta(I_D/I_H)$. Residues are coloured according to $\Delta(I_D/I_H)$ values. Residues with high $\Delta(I_D/I_H)$ are shown as cyan spheres, and residues with low $\Delta(I_D/I_H)$ are shown as red spheres. Residues which are unobservable in deuterated solvent in both 0 X_{SDS} and 0.769 X_{SDS} are shown as grey spheres.

3.3.5 Tryptophan side-chain indoles

Studying tryptophan residues by NMR provides invaluable information on both backbone amide and side-chain indole groups. Side-chains are exposed to different chemical environments compared to backbone amides (Figure 3.25a), and are directly involved in inter-helical interactions, hydrogen-bond networks and protein-retinal packing. Monitoring changes in tryptophan side-chain indoles by NMR therefore provides an extra probe for monitoring tertiary structure changes.

Like the backbone amides, pSRII tryptophan side-chain indole resonances showed complex chemical shift changes during unfolding mediated by different X_{SDS} (Figure 3.25b). Non-linear chemical shift trajectories and alternative resonances were observed, indicating the presence of different side-chain conformations exchanging at several different timescales.

Side-chain $\Delta\delta(^{1}H,^{15}N)$ correlate with the locations of the tryptophan residues (Figure 3.25c): Trp-76 and Trp-171 which pack directly against retinal show the smallest $\Delta\delta(^{1}H,^{15}N)$; the solvent-exposed Trp-60 and Trp-178 show bigger $\Delta\delta(^{1}H,^{15}N)$, possibly due to small changes in the solution phase with increasing X_{SDS}; Trp-9 and Trp-24, both of which face the detergent micelle, experience the biggest $\Delta\delta(^{1}H,^{15}N)$ since SDS is incorporated into the micellar phase. The lack of obvious correlation of side-chain $\Delta\delta(^{1}H,^{15}N)$ with SDS-mediated unfolding could reflect that the side-chains are generally maintained in very similar chemical environments.

Decreased peak intensities were observed with increasing X_{SDS} for most tryptophan side-chain indoles (Figure 3.25d). Consistent decreases in peak intensities were observed, indicating that with increasing X_{SDS}, side-chain motions are becoming increasingly dominated by µs–ms timescale conformational motions. Only Trp-24 side-chain indole initially became more flexible than it was in 0 X_{SDS}. Since Trp-24 is located around one helix turn away from the cytoplasmic end of helix A, this observation could reflect fraying at the end of helix A. Beyond 0.834 X_{SDS}, the alternative conformations of Trp-9 and Trp-24 side-chain indoles showed increasing flexibility with further increases in X_{SDS}. This highlights the heterogeneous nature of the unfolded ensemble.

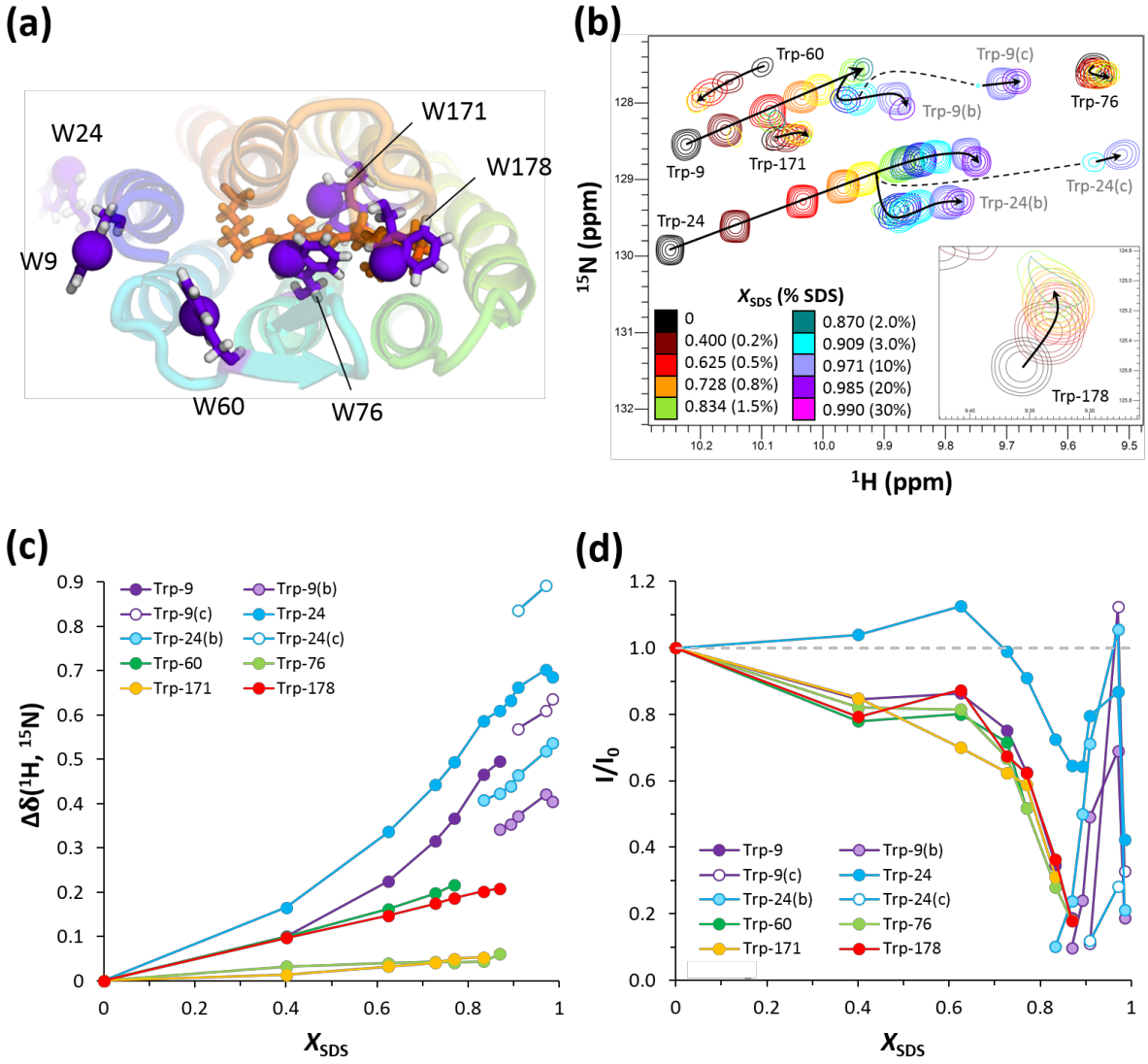


Figure 3.25: 2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC of tryptophan side-chain indoles
(a) Tryptophan side-chains are shown as sticks, with the indole N atoms shown as spheres. **(b)** Expanded view of the tryptophan side-chain indole region from an overlay of 2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC spectra recorded in different X_{SDS}. Changes in peak shifts are indicated by black arrows, with large shift changes indicated using dashed lines. Inset shows the Trp-178 resonance. Alternative conformations (labelled as b and c) are observed for Trp-9 and Trp-24 side-chain indoles. **(c–d)** X_{SDS}-dependence of $\Delta\delta(^{1}H,^{15}N)$ **(c)** and I/I_0 **(d)** for tryptophan side-chain indoles.

H/D exchange data are displayed as bar charts showing I_D/I_H for tryptophan side-chain resonances in 0 X_{SDS} and 0.769 X_{SDS} (Figure 3.26). Only one resonance is shown for each tryptophan, as alternative conformations have not appeared at 0.769 X_{SDS}. Statistical analysis could not be done for side-chain studies, since there are only six tryptophan side-chain resonances. A threshold of 5% difference in I_D/I_H between 0 X_{SDS} and 0.769 X_{SDS} was taken

to identify residues with important changes in solvent accessibility. Since tryptophan fluorescence is mediated by induced dipole moment and redistribution of electron density across the indole ring, changes in tryptophan fluorescence might be correlated with changes in the local environment of the side-chain detected by H/D exchange. H/D exchange measurements on tryptophan side-chain indoles were therefore compared against those of fluorescence intensity to understand the origins of tertiary structure changes detected by tryptophan fluorescence.

In 0 X_{SDS}, Trp-76 and Trp-171, both of which pack directly against the retinal polyene chain, incorporate less deuterium than other tryptophan side-chains. Out of the six tryptophan residues, only Trp-76 showed a significant increase in deuterium incorporation and solvent accessibility in 0.769 X_{SDS}. Since changes in tryptophan fluorescence are dominated by unquenching of tryptophan fluorescence due to retinal liberation and increase in solvent accessibility due to denaturation, it can therefore be deduced that Trp-76 is likely to have the greatest contribution towards the increases in fluorescence intensity and λ_{max} upon SDS-mediated unfolding as detailed in Chapter 2. Less H/D exchange was observed for Trp-171 side-chain indole in 0.769 X_{SDS} than in 0 X_{SDS}, suggesting lower solvent accessibility in the presence of SDS. The other tryptophan side-chain indoles, all of which either face the detergent micelle or are located at solvent-exposed regions, experience little changes in solvent accessibility. Altogether, the results from H/D exchange suggest that changes in tryptophan fluorescence (see Chapter 2) are likely dominated by changes at Trp-76, with a smaller contribution from Trp-171, and are thus reflective of the direct packing of Trp-76 and Trp-171 against the retinal chromophore in the folded state of pSRII.

Summarising the above analyses, tryptophan side-chains generally experience little changes in chemical environment, consistent with the protein remaining embedded in a micellar environment. Only side-chains which are facing the detergent micelle experience bigger changes in chemical environment due to incorporation of SDS into the micelles. Side-chain motions become increasingly dominated by µs–ms timescale motions in increasing X_{SDS}. Side-chains which pack against the retinal (Trp-76 and Trp-171) show the biggest changes in solvent accessibility, and are likely to also dominate changes in tryptophan fluorescence (see Chapter 2).

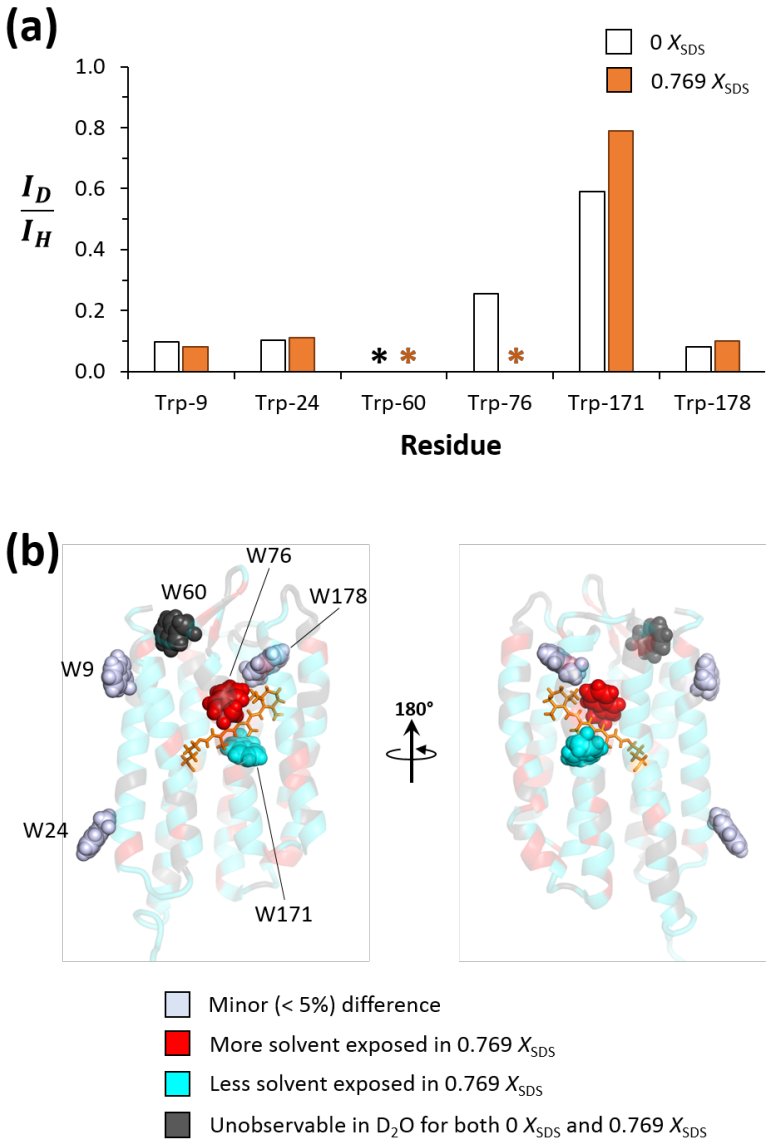


Figure 3.26: H/D exchange of tryptophan side-chain indoles
(a) Deuterium incorporation at each tryptophan side-chain indole is quantified by I_D/I_H. Residues which are unobservable are represented using asterisks (*). **(b)** Tryptophan side-chains are shown as spheres on the pSRII structure and coloured according to differences in I_D/I_H in 0 X_{SDS} vs. 0.769 X_{SDS}.

3.3.6 Backbone dynamics of pSRII in SDS/c7-DHPC mixed micelles

SDS unfolding of pSRII leads to changes in backbone dynamics, with some residues even experiencing motions at different timescales in different X_{SDS} (see Section 3.3.3). ^{15}N Relaxation NMR experiments were recorded to determine the timescales at which different motions occur. pSRII unfolding begins to occur in > 0.700 X_{SDS}, as determined by

increase in the unfolded fraction and λ_{max} of dead-time UV/vis spectra (see Chapter 2). Given the long timeframes of relaxation experiments (up to weeks), a low X_{SDS} of 0.728 X_{SDS} was used to avoid potential complications with aggregation and mixtures of unfolded and folded species.

^{15}N R_1 and $R_{1\rho}$ relaxation dispersion experiments were recorded for pSRII in 0 X_{SDS} (sample preparation and NMR experiment performed by Dr. Mark Bostock) and 0.728 X_{SDS} on a 600 MHz spectrometer, corresponding to a magnetic field strength of 14.1 T. R_2 values were extracted from $R_{1\rho}$ values using Equation 3.42. The relaxation rate constants for each assignable residue are plotted in Figure 3.27. High ^{15}N R_1 or low ^{15}N R_2 are indicative of increased motions on the fast sub-ns timescales typically related to a reduction in the order parameter S^2. Consistent with previous studies [337], most residues in the structured transmembrane regions of the protein show an R_1 of around 0.47 Hz, indicative of relaxation which is dominated by the overall molecular tumbling motion. Residues with high R_1 are identified by taking a cut-off at one standard deviation above the mean across helical residues. These residues are populated at the N-terminus, helix ends, loops AB, BC, and EF, and the C-terminal tail. Two residues located at the solvent/micelle interface, namely Gly-12 on helix A and Gly-136 on helix E, also have high R_1 (Figure 3.28). A similar distribution of residues is found for residues with low R_2, with the cut-off taken as one standard deviation below the mean across helical residues (Figure 3.28). Hence, based on R_1 and R_2, most of the transmembrane region backbone amide moieties are well-ordered with relaxation dominated by overall tumbling, while increased motional amplitude related to faster ps–ns motion is detected at helix ends and solvent exposed regions of the protein.

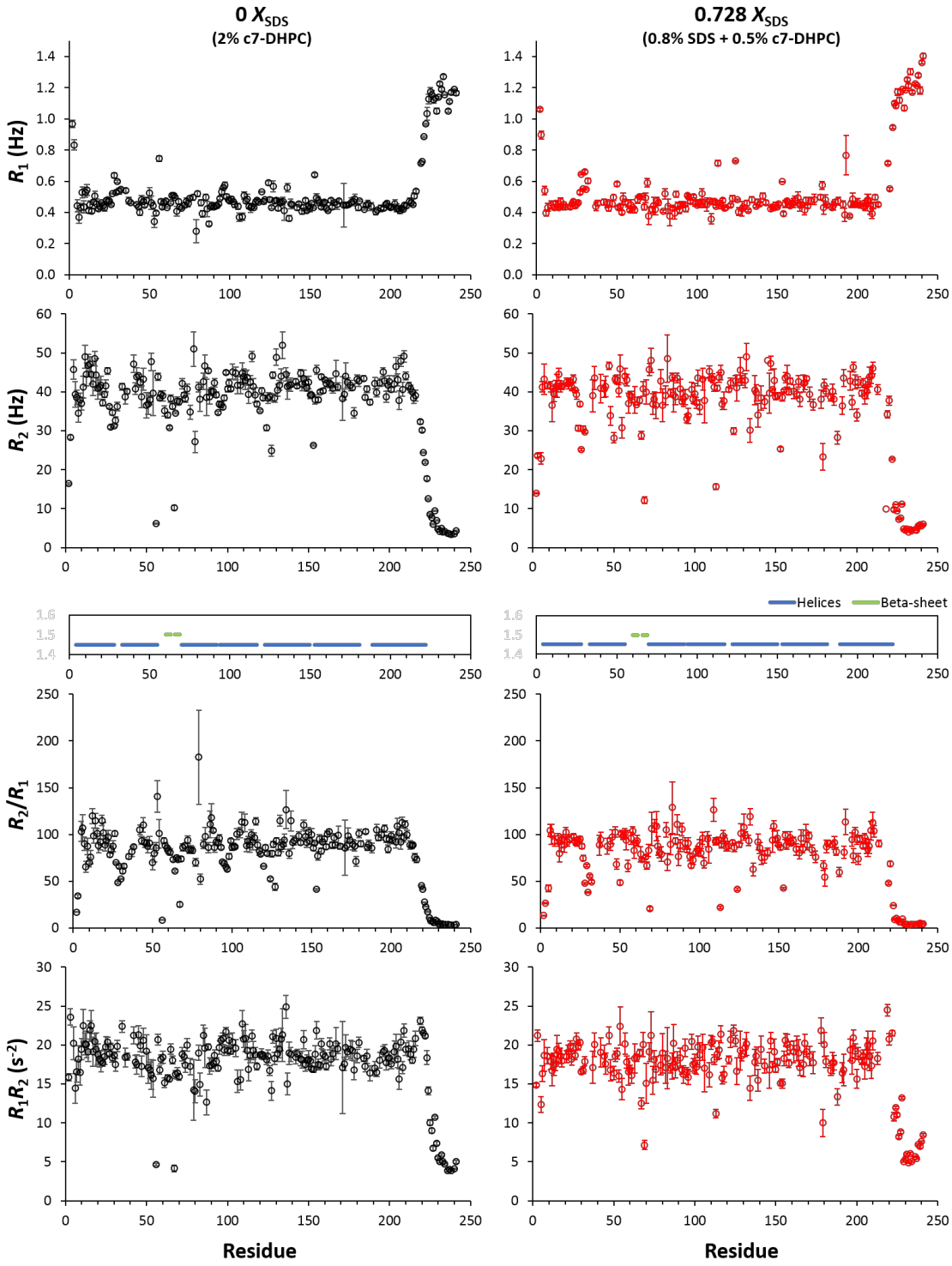


Figure 3.27: R_1 and R_2 relaxation rate constants for ^{15}N-pSRII in 0 X_{SDS} and 0.728 X_{SDS} measured at 14.1 T

R_1 and $R_{1\rho}$ experiments were recorded at 308 K on a 600 MHz spectrometer with a magnetic field strength of 14.1 T. R_2 values were extracted from $R_{1\rho}$ experiments. R_2/R_1 values were plotted to assess the overall rotational correlation time, and R_1R_2 values were determined to identify residues with R_{ex} contributions. Each sample contains ~ 525 µM ^{15}N-pSRII.

As shown in Equation 3.40, R_2/R_1 values are proportional to the rotational correlation time (τ_c). Average R_2/R_1 values taken across each helix were found to be similar within error (Table 3.2), indicating that the overall rotational correlation times of all helices remain similar regardless of the presence of SDS. This confirms that all helices remain structured and embedded in a micellar environment.

Secondary structure	Residues	0 X_{SDS} (2% c7-DHPC)		0.728 X_{SDS} (0.8% + 0.5% c7-DHPC)	
		R_2/R_1	τ_c (ns)	R_2/R_1	τ_c (ns)
Helix A	4–28	91 ± 17	30 ± 13	88 ± 16	30 ± 12
Helix B	33–55	93 ± 17	31 ± 13	87 ± 16	30 ± 13
Helix C	70–92	95 ± 26	31 ± 16	95 ± 15	31 ± 12
Helix D	94–117	89 ± 15	30 ± 12	84 ± 20	29 ± 14
Helix E	122–150	91 ± 18	30 ± 13	89 ± 17	30 ± 13
Helix F	153–181	89 ± 13	30 ± 11	85 ± 15	29 ± 12
Helix G	189–222	86 ± 25	30 ± 16	86 ± 21	30 ± 14
Strands in BC loop	60–63, 66–69	71 ± 21	27 ± 14	73 ± 25	27 ± 16

Table 3.2: Average R_2/R_1 and τ_c values taken across secondary structure elements in pSRII at 0 X_{SDS} and 0.728 X_{SDS}
Values are presented as the average ± standard deviation across the residues specified for each secondary structure element.

Variations in R_2/R_1 are obscured by the lack of distinction between effects of chemical exchange and anisotropic motion, whereas variations in R_1R_2 values are much less sensitive to effects from motional anisotropy [328]. The presence of fast ps–ns motions reduces R_1R_2; slower motions associated with R_{ex} increase R_1R_2. Calculations by Kneller *et al*. [328] show that for molecules with $\tau_c > \sim 5$ ns in the absence of R_{ex} at a magnetic field strength of 14.0 T, R_1R_2 is ~ 20 s^{-2}. 0.9 mM pSRII in 1.6% c7-DHPC at 35 °C gives τ_c = 25.7 ns, corresponding to a molecular weight of pSRII-c7 protein-detergent complex to be around 70–80 kDa [337]. The cut-off of 20 s^{-2} was therefore applied to identify backbone amides of pSRII which might have R_{ex} contributions. Given the strong spin-lock field used in the $R_{1\rho}$ experiment, the extracted R_2 values are therefore sensitive to R_{ex} contribution in the μs timescale.

In 0 X_{SDS}, residues with high R_1R_2 are found in solvent-exposed regions and at the micelle/protein interface. There also appears to be a cluster of transmembrane residues in the interface between helices A and B, on helix E facing helix D, suggesting possible µs motions indicative of helix-helix interactions. A few residues packing against the six membered ring of the retinal Schiff base, including Met-109, Gly-130 and Phe-134, and also Gly-207 and Gly-209 on helix G, show potential R_{ex} contributions (Figure 3.28). In the presence of SDS, very few residues with high R_1R_2 are in the transmembrane region. Some residues appear in the micelle/solvent interface at the extracellular side, including Trp-178 on the outer rim of the binding pocket (Figure 3.28). Notably residues in the interface between helices A and B no longer have high R_1R_2, suggesting a possible smaller contribution of µs motion in this region in the presence of SDS. It is difficult to make conclusions about the region on helix E due to lack of assignments and/or reliable relaxation rate constants in this region.

The lack of µs timescale R_{ex} contributions in the transmembrane region in 0.728 X_{SDS} likely indicates that chemical exchange at these backbone amides has transitioned to the slower µs–ms timescale directly related to line broadening, as evidenced by decreasing peak intensities in higher X_{SDS}. Since the strong spin-lock field used in $R_{1\rho}$ experiments only permits the investigation of low/sub-µs timescale motions, other motional regimes would not be accessible in this study. Further confirmation for the presence of ms–µs timescale motions can be done using CPMG relaxation dispersion experiments. R_{ex} contributions of different frequencies are removed by altering the frequency at which the 180° refocusing pulses are applied (ν_{CPMG}). The size of R_{ex} contribution towards R_2 at each backbone amide is obtained by plotting the observed R_2 against ν_{CPMG}. Once R_2 values are corrected for any R_{ex} contribution, R_1, R_2 and NOE measurements can then be used for characterising internal molecular motions on the sub-nanosecond timescale by calculating order parameters using the Lipari-Szabo formalism.

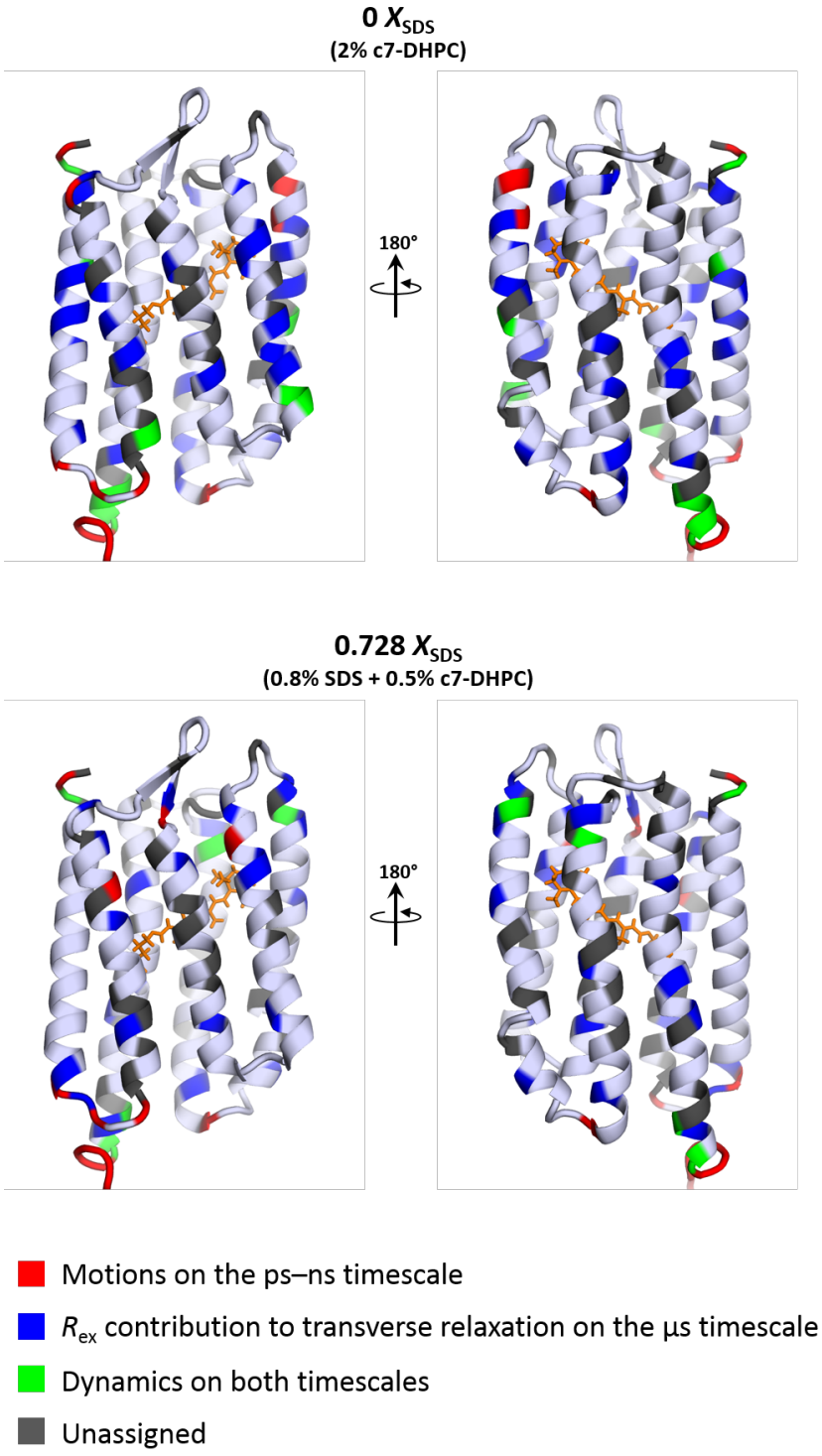


Figure 3.28: Residues experiencing backbone dynamics on different timescales
Residues with motions on the ps–ns timescale are identified as those having high ^{15}N R_1 of at least one standard deviation above the average across all helical residues. Residues with potential R_{ex} contributions to ^{15}N R_2 are identified by taking a threshold of $R_1R_2 > 20$ s^{-2} [328].

3.4 Results: Loss of retinal chromophore alters the backbone dynamics of transmembrane helices

Hydroxylamine-mediated Schiff base hydrolysis in the absence of SDS takes several hours to reach completion. Light illumination accelerates hydroxylamine-mediated Schiff base hydrolysis [304]. To obtain sensory opsin apo-protein, pSRII was illuminated for 1 h using a 524-nm green LED (300 μmol m^{-2} s^{-1}) in the presence of 100 mM hydroxylamine at 4 °C. Hydroxylamine was then removed by buffer exchange, and the resulting sample concentrated and studied by NMR.

Hydroxylamine-treated sensory opsin showed a large loss in helical content (Figure 3.29a), supporting the proposed role of retinal in pSRII secondary structure formation (see Chapter 2). 1D ^{19}F NMR on ^{19}F-5-Trp-labelled pSRII shows one single broad peak, suggesting several tryptophan side-chains are experiencing a similar chemical environment and are undergoing conformational exchange in the μs–ms timescale (Figure 3.29c). ^{1}H NMR of sensory opsin showed several strong peaks in the 7.0–8.5 ppm region while peaks in other regions had very low intensities (Figure 3.29e). These observations further suggest a key role of retinal in the structure and dynamics of backbone amides. Irreversible aggregation is unlikely within the short timeframe of the experiments (within 12 h). Addition of 0.943 X_{SDS} (1% SDS in 0.1% c7-DHPC) to sensory opsin restores the α-helical content to similar levels observed in pSRII unfolded in 0.769 X_{SDS} without hydroxylamine (Figure 3.29b). However, ^{19}F NMR of sensory opsin in 0.834 X_{SDS} shows one single broad peak, albeit with narrower apparent linewidth than that observed for sensory opsin without SDS (Figure 3.29d). Likewise, the ^{1}H NMR profile for sensory opsin in 0.834 X_{SDS} was different compared to that for pSRII unfolded using 0.834 X_{SDS} (Figure 3.29f), indicating structural and dynamic differences between hydroxylamine-treated sensory opsin vs. SDS-unfolded pSRII despite both samples harbouring a hydrolysed Schiff base and solubilised in SDS/c7-DHPC micelles. This indicates that although the total secondary structure content appears to be similar for SDS-unfolded pSRII and hydroxylamine-treated sensory opsin in 0.834 X_{SDS}, the structural properties and dynamics of tryptophan side-chains and backbone amides in the respective samples are different. Further conclusions could not be drawn from 2D NMR spectra either

because of significant peak broadening. The observed resonances likely correspond to the residues in the flexible C-terminal tail (Figure 3.29g–h).

Altogether, these results suggest that the retinal chromophore underpins the structural and dynamic properties of pSRII. Disruption of native protein-retinal interaction by Schiff base hydrolysis leads to μs–ms timescale motions across the protein, possibly suggesting the uncoupling of the seven helices from each other. Studies using other techniques in which data quality is not affected by μs–ms timescale motion will be helpful for understanding the role of retinal in the folding pathway and discerning structural differences in sensory opsin apo-protein obtained from hydroxylamine-mediated Schiff base hydrolysis vs. SDS-mediated protein unfolding.

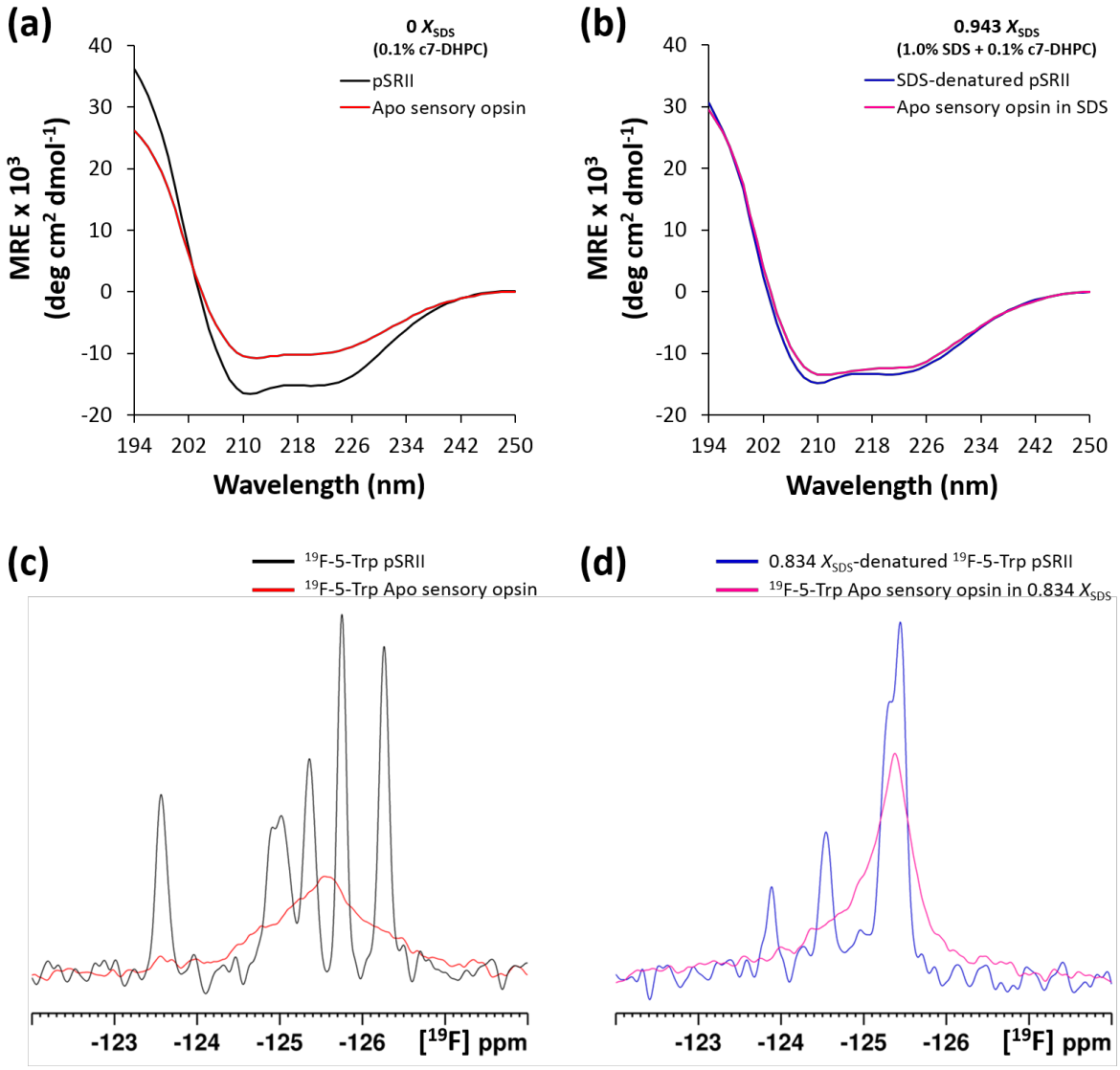


Figure 3.29: Structural characterisation of apo sensory opsin in different X_{SDS}
Circular dichroism **(a–b)** and ^{19}F NMR **(c–d)** of apo sensory opsin are overlaid with those measured for pSRII in different X_{SDS}.

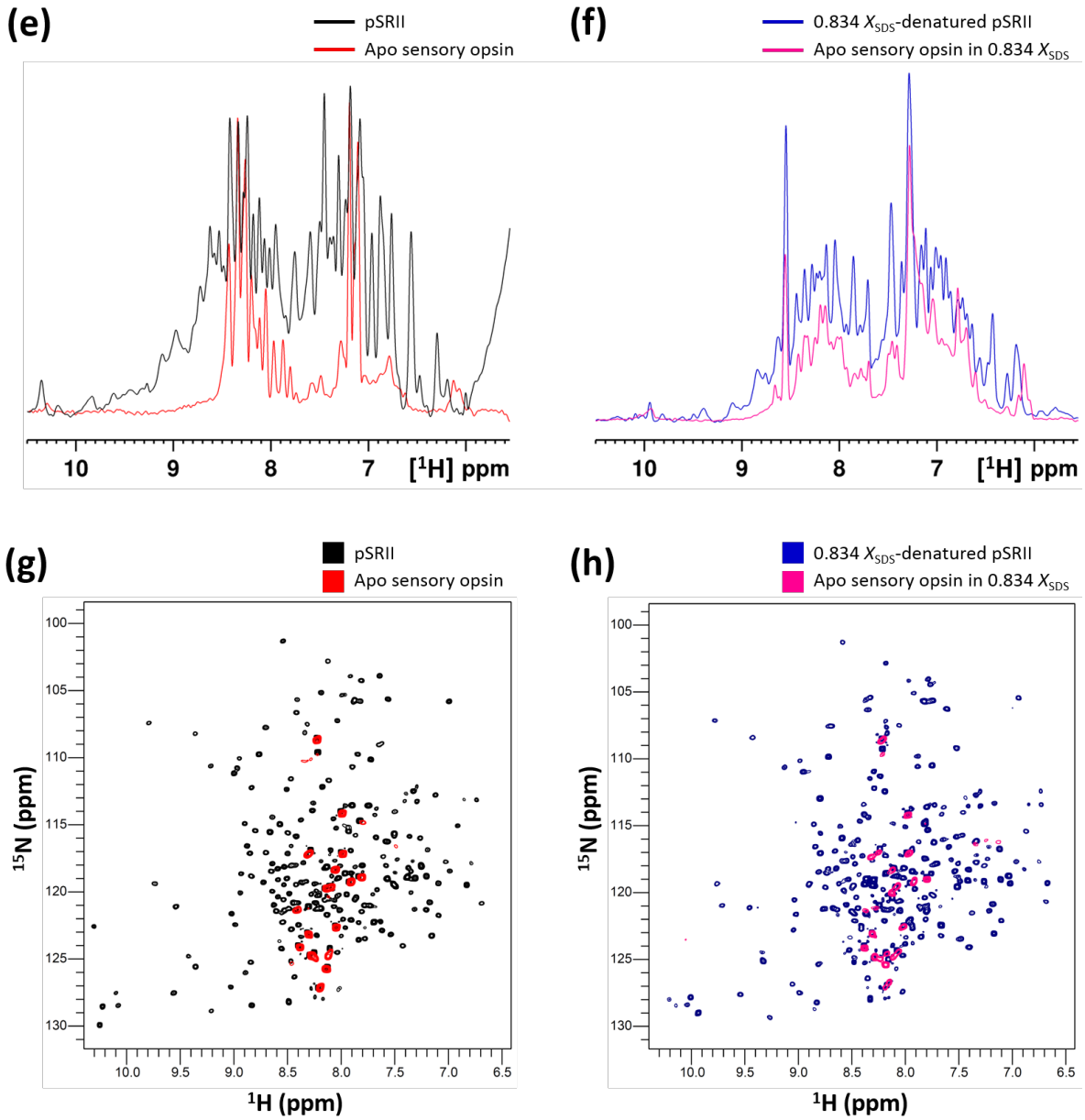


Figure 3.29: Structural characterisation of apo sensory opsin in different X_{SDS} (cont.)
^{1}H NMR **(e–f)** and 2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC **(g–h)** of 162 µM apo sensory opsin are overlaid with those measured for 90 µM pSRII in different X_{SDS}.

3.5 Results: Structural properties of acid-denatured pSRII

Denaturant screening and biophysics studies (see Chapter 2) showed that SDS at pH 2.0 led to rapid unfolding of pSRII within seconds followed by very slow Schiff base hydrolysis occurring over the course of several days. This enabled prolonged observation of the SR_{440} species, which is characteristic of protein with retinal remaining covalently attached via the protonated Schiff base but devoid of native retinal-protein packing. Taking into consideration the previously proposed role of retinal in stabilising pSRII (see Chapter 2 and Section 3.4), pSRII at pH 2.0 is an interesting candidate for NMR studies to further understand the structural properties of SR_{440} in relation to folded and SDS-denatured pSRII.

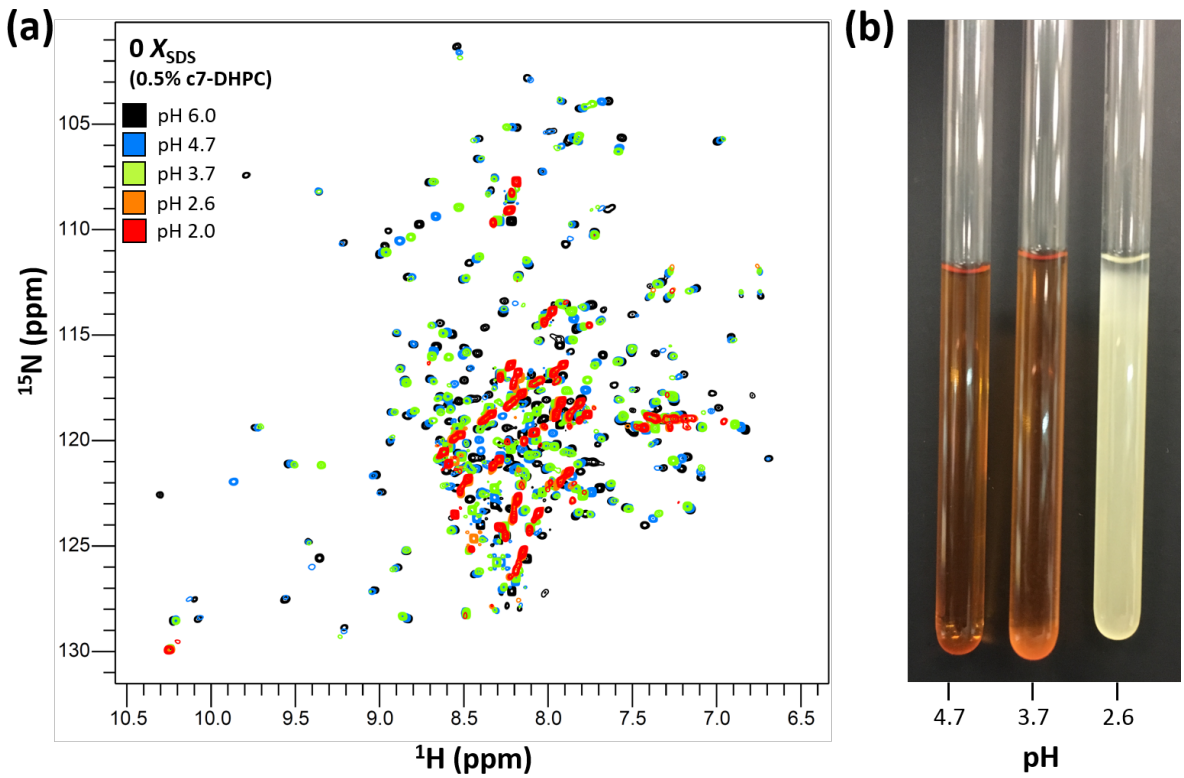


Figure 3.30: Acid denaturation of pSRII in 0 X_{SDS}
(a) 2D ^{1}H,^{15}N TROSY-HSQC spectra of pSRII at different acidic pHs. All spectra were recorded on a 800 MHz spectrometer at 308 K with 240 scans. All samples contained ~ 90 µM ^{15}N-pSRII and TSP as internal reference. **(b)** Photos of ~ 90 µM ^{15}N-pSRII at different acidic pHs after one week at room temperature.

2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC spectra were recorded for pSRII in progressively lower pHs (pH 6.0, pH 4.7, pH 3.7, pH 2.6, pH 2.0) in the presence of 0.5% c7-DHPC (no SDS) (Figure 3.30a). $\Delta\delta(^{1}H,^{15}N)$ were generally small (Figure 3.31a), with $\Delta\delta(^{1}H,^{15}N)$ < 0.1 ppm at pH 4.7 and pH 3.7 for most transmembrane residues. Larger $\Delta\delta(^{1}H,^{15}N)$ were observed across all residues at lower pH. Only two residues within the binding pocket had $\Delta\delta(^{1}H,^{15}N)$ > 0.1 ppm: Thr-80 on helix C (pH 4.7 and pH 3.7) and Met-15 on helix A (pH 3.7) (Figure 3.31c). This indicates that mildly acidic pHs led to minimal disruption of the binding pocket, consistent with the samples remaining orange in colour after one week at room temperature (Figure 3.30b). Several residues which span three helical turns at the cytoplasmic end of helix D facing helix E showed $\Delta\delta(^{1}H,^{15}N)$ > 0.1 ppm at pH 4.7, while $\Delta\delta(^{1}H,^{15}N)$ > 0.1 ppm in other helices are within one turn from helix ends (Figure 3.31b), and helix ends continued to show greater $\Delta\delta(^{1}H,^{15}N)$ than transmembrane residues.

Acid denaturation also led to line broadening, hence residues become increasingly unobservable (Figure 3.32a). Such residues were easily identified especially since $\Delta\delta(^{1}H,^{15}N)$ were small across the protein. Residues which became unobservable at pH 4.7 are located in the extracellular half, of which Trp-76, Thr-78, Met-109 and Trp-178 are residues within the binding pocket (Figure 3.32b). An extensive stretch of helix C and Val-206 and Gly-207 on helix G are also noted to have become unobservable, suggesting increased μs–ms timescale motion for residues facing the retinal binding pocket and residues near the Schiff base. At pH 3.7, almost the entire helix C became unobservable, suggesting concerted changes in amide backbone dynamics across the entire helix. More residues within the binding pocket disappeared, including Ala-47 on helix B, Ile-83 on helix C, Gly-112 on helix D, Thr-204, Lys-205 and Phe-208 on helix G, indicating further loosening of the binding pocket and the Schiff base attachment site. Interestingly, many residues which are unobservable at pH 4.7 or pH 3.7 are hydrogen-bonded to the internal water network in the crystal structure of pSRII (PDB 1H68; [336]) (Figure 3.32c). This suggests that acid denaturation leads to disruption of the internal water network. The abolishment of key hydrogen-bonds uncouples the key residues from each other, enabling the backbone amides to sample different conformations on a μs–ms timescale. At pH 2.6 and pH 2.0, most helical residues became unobservable, indicating altered backbone dynamics across the entire protein. This is noted to be correlated with the samples appearing pale yellow (complete Schiff base hydrolysis) and pronounced

precipitation over several days (Figure 3.30b), suggesting extensive exposure of aggregation-prone surfaces and/or decreased protein stability.

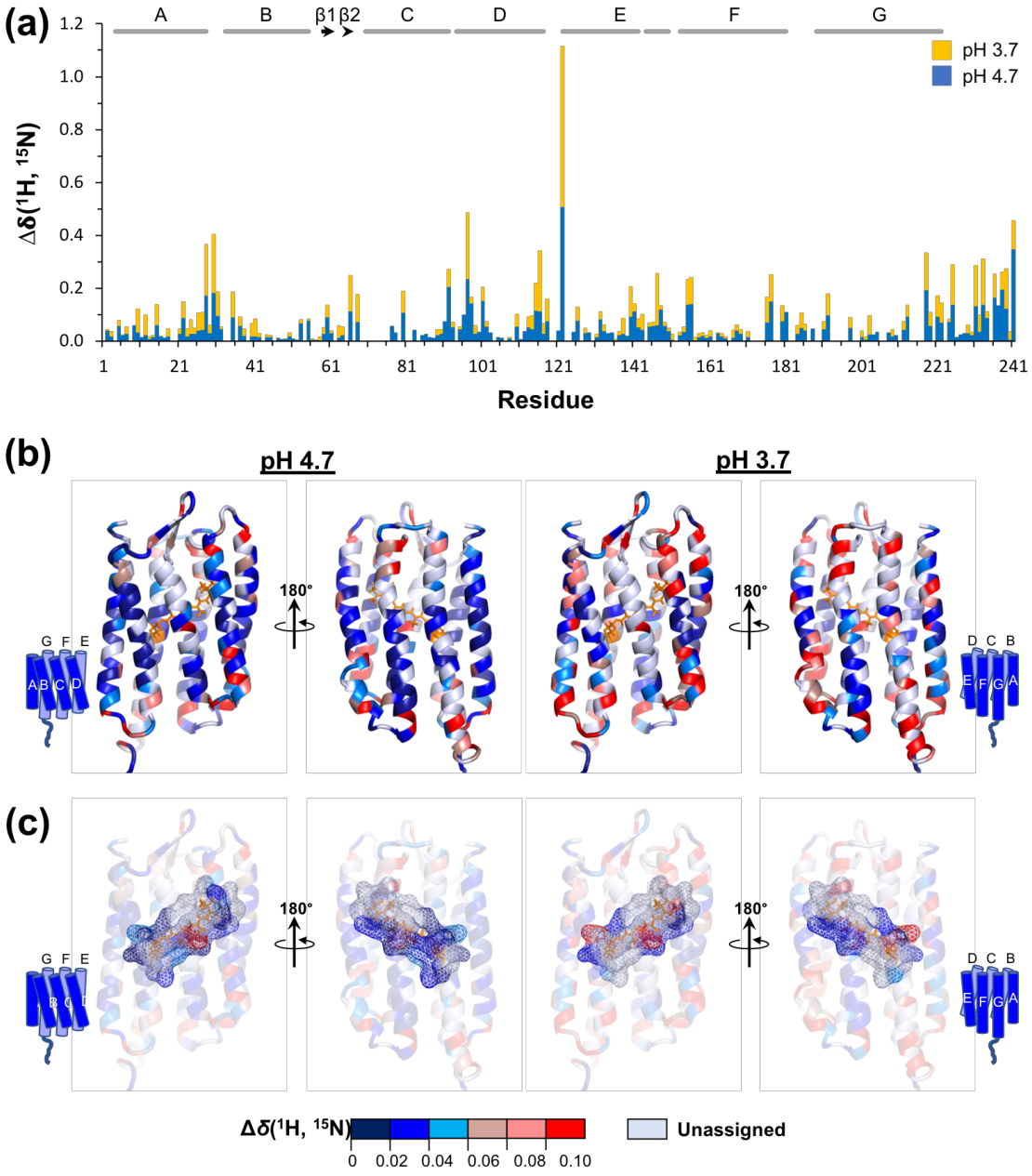


Figure 3.31: Chemical shift perturbation of acid-denatured pSRII relative to folded pSRII at pH 6.0

(a) Combined chemical shift differences ($\Delta\delta(^1H,^{15}N)$) of pSRII at pH 4.7 and pH 3.7. Helices are shown as grey bars and β-strands as black arrows. **(b)** $\Delta\delta(^1H,^{15}N)$ are mapped onto the pSRII structure. Residues are coloured from blue to red according to $\Delta\delta(^1H,^{15}N)$ values, showing greater structural changes at the ends of helices than in the transmembrane region. Unassigned residues are coloured white. **(c)** Surface representation of residues within 3 Å of retinal coloured according to $\Delta\delta(^1H,^{15}N)$ values to show structural changes at the retinal binding pocket.

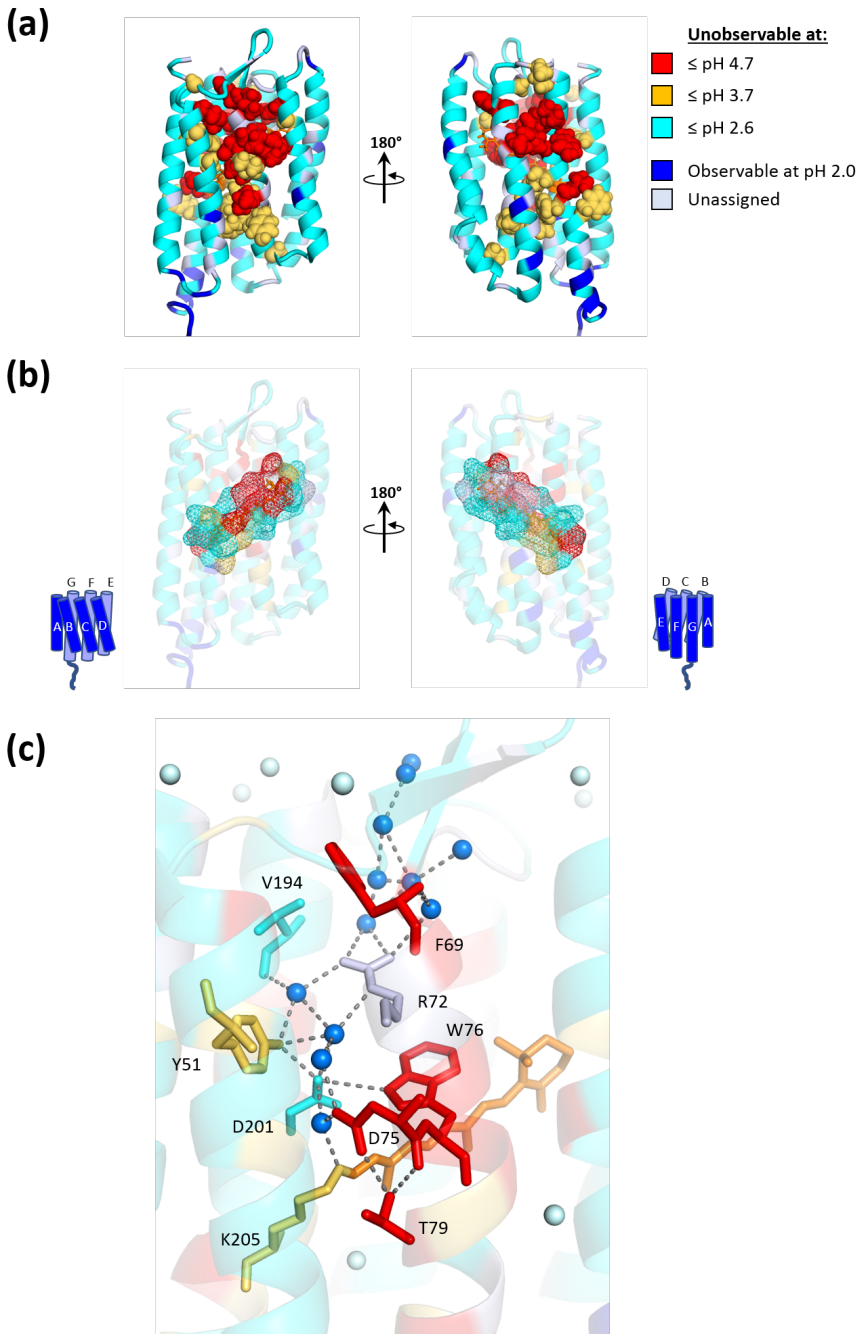


Figure 3.32: Residues which are unobservable at different pHs mapped onto pSRII structure
(a) Residues which become unobservable at pH 4.7 and pH 3.7 are shown as red and yellow spheres, respectively. Residues which become unobservable at pH 2.6 are coloured cyan. Residues which remain observable at pH 2.0 are coloured blue. **(b)** Surface representation of residues within 3 Å of retinal coloured according to the pH at which the residues become unobservable, showing changes in pSRII backbone dynamics at the retinal binding pocket. **(c)** Residues which are hydrogen-bonded to the internal water network are shown as sticks on the crystal structure of pSRII (PDB 1H68) and coloured according to the pH at which the residue becomes unobservable. The hydrogen-bonds in the internal water network are shown as grey dashed lines. Water molecules are shown as spheres, with internal water molecules coloured blue.

3.6 Discussion

High-resolution studies on intact unfolded membrane proteins are scarce in the literature. In view of NMR studies, this likely reflects the challenges in studying a heterogeneous ensemble of unfolded structures with extensive μs–ms timescale motions, and obtaining uniformly isotopically-labelled membrane protein samples. This chapter presents detailed NMR studies on backbone amides and tryptophan side-chain indoles of pSRII under different conditions (SDS-unfolded, acid-denatured, and retinal Schiff base hydrolysed by hydroxylamine). Combined analyses of chemical shifts, peak intensities and H/D exchange were performed to investigate structural and dynamics changes.

3.6.1 Unfolding mechanism of pSRII

pSRII has a rotational correlation time of ~30 ns, and the transmembrane helices are well-ordered. Unfolding pSRII in SDS or at acidic pHs along with hydroxylamine-mediated retinal Schiff base hydrolysis reveals a backbone amide NMR fingerprint that is dominated by μs–ms timescale conformational exchange dynamics.

For pSRII denatured in increasing X_{SDS}, loosening of the binding pocket correlates with increasing μs–ms timescale motion in backbone amides beginning at the cytoplasmic interface of helices C, E, F and G and spreading across the protein, while the extracellular half of helix D adopts faster motion in high X_{SDS}. Changes in chemical shifts are generally small, but clearly show that disruption in the packing of helices C and D against the retinal chromophore begins before the unfolding transition and is followed by small changes in helix G near the site of retinal attachment beyond the unfolding transition. Out of the six tryptophan residues, an increase in solvent accessibility was observed at Trp-76. Transmembrane helices in DDM-reconstituted bacteriorhodopsin have a rotational correlation time of ~35 ns [338], similar to that of pSRII. A previous study of the backbone dynamics of a proteolytic fragment containing residues 1–71 of bacteriorhodopsin reconstituted in SDS micelles shows that backbone NH vectors are involved in two types of internal motion: fast motion (< 20 ps), which is attributed to thermal fluctuations of atoms; and intermediate motion (~ 1 ns), which is attributed to helix bending. In addition, most backbone amide groups are involved in co-operative exchange processes over the rate range

10^3–10^4 s^{-1}, indicative of dynamic helix-helix interactions, such as helix tilting, the helices sliding past each other, extending or shortening [331]. Such µs–ms timescale motions were also observed for many backbone amides of pSRII unfolded in SDS/c7-DHPC mixed micelles, suggesting SDS-unfolded pSRII might be experiencing broadly similar dynamics as the bacterioopsin fragment. Further confirmation of the timescales of pSRII NH bond vector internal motions can be derived from order parameters if heteronuclear NOE experiments were recorded. 2D [^{1}H,^{15}N] HSQC spectra of folded [α-^{15}N]-Lys-labelled rhodopsin in DDM suggests that µs–ms timescale motions are observed for most lysine backbone amides except for a C-terminal lysine which shows motion on the ns timescale. Rhodopsin has multiple backbone conformations, as more resonances with variable intensities were observed than the number of lysines in [α-^{15}N]-Lys-labelled rhodopsin [339]. In contrast, each tryptophan indole side-chain appears to be restricted to one conformation [340]. SDS denaturation leads to increased flexibility of backbone amides in unfolded regions, as evidenced by an increase in the number of observable signals in 30% SDS. Electron paramagnetic resonance (EPR) spectroscopy experiments show that the extracellular domain retains more rigidity than cytoplasmic residues in SDS-denatured rhodopsin, consistent with the proposed location of residual structure in denatured states of rhodopsin [138].

To summarise, SDS denaturation leads to changes in protein backbone flexibility across the three proteins. Backbone amides of folded pSRII are well-ordered, with relaxation dominated by overall tumbling. Widespread µs–ms timescale conformational dynamics are observed during the unfolding of pSRII and bacteriorhodopsin due to changes in helix-helix interactions. In comparison, folded dark-state rhodopsin experiences slow µs–ms timescale motions, and SDS leads to increases in flexibility and faster ns timescale dynamics in unfolded regions while residues constituting the residual structure in the extracellular domain retain more rigidity. Once again, the NMR studies presented in this chapter are consistent with the corresponding biophysics and kinetic studies, and point towards differences in the unfolding mechanisms and protein-SDS interactions amongst the three proteins: SDS-denatured bacteriorhodopsin and pSRII undergo mainly changes in tertiary structure (helix packing) whereas SDS-denatured rhodopsin shows changes in both secondary and tertiary structures.

For acid-denatured pSRII, changes in chemical shifts were very small in the transmembrane region, consistent with the protein remaining embedded in a micellar environment and little changes in α-helical content even at pH 2.0. Various residues which are hydrogen-bonded to the internal water network exhibit μs–ms timescale motion, suggesting disruption of the retinal binding pocket and the internal water network. The internal water network integrates the seven helices into a tight bundle. Crystal structures of ground state and the light-activated M-state pSRII show that signal transduction during light activation is mediated by protonation of Asp-75 and charge rearrangement of the active site, leading to breaking of hydrogen bonds in the internal water network and loss of connectivity between helices C and G [341]. It is interesting to speculate whether the acid denatured state could be somewhat similar to the light-activated state. The pKa of Asp-75 is 3.4 for pSRII in phosphatidylcholine and DDM, and different membrane mimetics could lead to slight changes in pKa [342]. The acidic pHs used in this study could lead to light-independent protonation of Asp-75, followed by charge rearrangement of the active site and breaking of the internal water network in a comparable manner to light activation. Disruption of the internal water network could be further evidenced by the observation of μs–ms timescale motion across almost the entire helix C at pH 3.7, suggesting that helix C is the first to be uncoupled from the helix bundle, somewhat analogous to the loss of connectivity between helices C and G in the light-activated state. Further lowering of the pH leads to μs–ms timescale motion across the entire transmembrane region, indicating widespread changes in helix-helix interactions.

Large changes in backbone dynamics were observed when retinal was removed by hydroxylamine, such that only one single broad peak was observed in 1D ^{19}F NMR spectra, both in the absence and presence of SDS. This indicates that the retinal chromophore exerts a key role in restricting the movements of the seven transmembrane helices, and liberation of the retinal chromophore enables more sliding movement of the helices to occur on a μs–ms timescale. However, further conclusions could not be drawn due to poor spectral quality. Bacterioopsin (bO) produced from a retinal-deficient E1001 *Halobacterium salinarum* strain gave rise to very broad ^{13}C solid-state NMR signals at 20 °C, whereas bO from hydroxylamine-bleached bacteriorhodopsin shows similar spectral features to folded bacteriorhodopsin except for lower peak intensities [343]. It is speculated that the trimeric conformation is retained in hydroxylamine-bleached bO even after retinal is removed,

enabling stabilisation by hydrophobic packing through helix-helix interactions between neighbouring protomers and hence certain secondary structures and backbone dynamics are retained [343]. In contrast, bO from the retinal-deficient strain has more similar spectral properties to hydroxylamine-bleached pSRII. While it remains unclear how the backbone conformation of bacteriorhodopsin is altered upon Schiff base formation, it has been speculated that bO from the retinal-deficient strain represents a partially folded apo-protein undergoing slow chemical exchange among various slightly different conformers on the ms timescale, and these slow motions are speculated to be essential for entry of retinal into bO [343]. Dark-state rhodopsin does not react with hydroxylamine [314], and the retinal becomes accessible to hydroxylamine only upon formation of the Meta II state. Hence, comparison on the effect of retinal across the three proteins can only be done using light-activated rhodopsin. Light activation of rhodopsin triggers isomerisation and movement of the retinal on the ps–ns timescale. These fast motions are proposed to initiate collective fluctuations of helices H5 and H6 and cytoplasmic loops in the Meta I-Meta II equilibrium on the μs–ms timescale, priming the receptor for G protein activation [344]. It is noted that more transmembrane helices in bacterioopsin and sensory opsin experience movements on the μs–ms timescale than in bovine opsin. This could perhaps reflect functional differences between the photocycles of microbial vs. mammalian rhodopsins – the retinal chromophore remains covalently attached during the photocycles of microbial rhodopsins, whereas retinal dissociates from mammalian rhodopsins and the chromophore is subsequently regenerated. It might therefore be inferred that liberation of the retinal chromophore from microbial rhodopsins could lead to more prominent changes in protein structure and dynamics. It is tempting to speculate that the retinal chromophore might influence the protein folding energy landscape by perturbing the equilibrium amongst different backbone conformations of the apo-protein state. Formation of the Schiff base (covalent linkage to Lys-205) and other native retinal-protein interactions during protein folding could, for example, stabilise particular conformations to favour folding and generation of the native chromophore. Hypotheses on the possible influence of the retinal chromophore on protein folding has already emerged from studies on bovine rhodopsin by Sakamoto and Khorana [319], and the kinetics and the extent of retinal-binding during the regeneration of bacteriorhodopsin from bacterioopsin have been studied using UV/vis spectroscopy [87,107,133,317,345], circular dichroism [107,132], tryptophan fluorescence [120,317,345] and calorimetry [346].

To summarise, retinal plays a key role in restricting helix movements in pSRII, bacteriorhodopsin and bovine rhodopsin. Liberation of the retinal enables the protein to sample different conformations through helix movements on the μs–ms timescale. These motions could be important for retinal binding and downstream biological functions (light activation or signalling). More detailed studies on the structure and dynamics of the apo-protein are expected to yield further insights on the roles of the retinal chromophore in the folding of pSRII and other retinal-binding proteins, and potentially open new avenues for investigating the importance of ligand-binding in the folding of other ligand-binding membrane proteins (e.g. GPCRs).

Summarising the NMR studies in this chapter, SDS-denaturation, acid-denaturation and hydroxylamine-bleaching all led to widespread observation of slow μs–ms timescale motions across pSRII backbone amides, as constraints on protein backbone motions originally imposed by retinal become relaxed upon liberation of the chromophore. Precisely which retinal-protein interactions are being disrupted and the final structural state would depend on the denaturation/solvent condition employed. This conclusion is most clearly exemplified by the different distribution of unobservable residues in acid-denatured vs. SDS-denatured pSRII, suggesting different unfolding pathways. A similar conclusion has been obtained from mapping the unfolding processes of bovine α-lactalbumin in SDS, acid, GdmCl, TFE and heat by 1D ^{1}H NMR combined with principal component analysis (GPS-NMR), showing that these denaturants move the protein away from the folded state in different directions [347], hence indicating different unfolding pathways.

3.6.2 Challenges and limitations of analysis

The NMR studies presented in this chapter have provided considerable insights into structural and dynamics changes to pSRII denatured by different methods. NMR, especially the ^{15}N relaxation experiments, were originally envisioned to provide information on the unfolding pathway by verifying whether individual helices or groups of helices constitute the folding core, as individual helices or helical segments which retain residual structure are expected have different backbone dynamics compared to other unfolded regions. However, a number of challenges prevent clear interpretation of the data in this light, mostly due to:

1) Small changes observed in the transmembrane region – this can be attributed to unfolded pSRII remaining embedded in detergent micelles, thus restricting tertiary structure changes in the transmembrane region, and most NMR analysis being done on backbone amides whereas tertiary structure changes are mediated by side-chains;
2) Resonances becoming unobservable at high X_{SDS} due to conformational exchange on the µs–ms timescale and possible complications with aggregation over prolonged experiment time.

Exchange processes taking place on the µs–ms timescale are particularly detrimental to spectral quality. This is because coherence transfer steps become inefficient in the presence of fast relaxation and the additional exchange contribution to transverse relaxation leads to very strong line broadening which very likely affects both correlated dimensions. Future studies could involve recording spectra at higher static magnetic fields to push the conformational exchange towards the slow exchange regime.

Simplified treatment of NMR titration experiments (e.g. only analysing chemical shift changes) could lead to incorrect mechanistic interpretations. Multi-state systems demonstrate unique features and ambiguous spectral features. Quantitative lineshape analysis must be performed for stringent identification of molecular mechanism. Considerable efforts have been devoted in NMR lineshape analysis for quantitative analysis of titration data based on the fitting of 1D spectra or cross-sections of 2D spectra [348]. Despite the wealth of information (binding model, interaction kinetics, lifetimes of different species) which could be offered by quantitative 2D NMR lineshape analysis, this option was not pursued as deriving rate constants for different backbone amides and obtaining a unified unfolding mechanism for pSRII would have been difficult. Apart from the aforementioned problems with small spectral changes, µs–ms dynamics and aggregation, it is uncertain whether 0.990 X_{SDS} (30% SDS + 0.5% c7-DHPC), the highest SDS concentration used in the titration series, represents the end-point of the titration. While the binding pocket is fully opened by 0.943 X_{SDS} (5% SDS + 0.5% c7-DHPC) (see Chapter 2), structural and dynamics changes might still be occurring at other parts of the protein in higher X_{SDS}. Incomplete titration series might distort interpretation of lineshape analysis.

Although the steady-state NMR methods employed in this chapter have enabled characterisation of equilibrium molecular dynamics occurring on a sub-second time scale, processes which are far-removed from equilibrium (e.g. protein unfolding in high X_{SDS}) are best studied by techniques with excellent time resolution. Real-time NMR combines rapid sample injection and mixing inside the magnet with fast data acquisition to record a series of NMR spectra during the kinetic reaction. Real-time NMR has been used to investigate the unfolding and folding pathways and conformational dynamics of small proteins. For example, SOFAST real-time 2D NMR studies on the backbone amides of α-lactalbumin confirm that a single transition state ensemble controls folding from the molten globule state to the native state [349]. Biophysics and kinetic studies in Chapter 2 indicate that native retinal-protein interactions are abolished within seconds to minutes, depending on X_{SDS} and pH. Studying changes in retinal-protein interactions at short unfolding times using real-time NMR would enable more advanced understanding on the sequence and timescale of structural and dynamics changes in SDS- and acid-denaturation.

3.6.3 Future work

Considering the challenges and limitations listed above, the results and interpretations derived from this chapter would be of greater value to the membrane protein field should these be successfully backed up by complementary experimental techniques in the future. In particular, experimental techniques which permit the study of μs–ms timescale motions would be particularly helpful in extending our knowledge on pSRII folding, considering the functional importance of helix motions on the μs–ms timescale.

3.6.3.1 Further characterisation of the structure and dynamics of denatured pSRII

It is clear from biophysics (see Chapter 2) and NMR studies that SDS-denaturation, acid-denaturation and hydroxylamine-bleaching lead to disruption of interhelical interactions and tertiary structure, yielding a heterogeneous ensemble of backbone amide conformations undergoing μs–ms timescale motions. More precise characterisation of structure and dynamics changes can be achieved using various different methods.

The heterogeneous ensemble of structures represented by denatured pSRII can be studied using double electron-electron resonance (DEER) spectroscopy [350], laser-induced

oxidative labelling of methionine side-chains coupled with mass spectrometry [315], or pulsed H/D exchange coupled with mass spectrometry (HDX-MS) [351]. These techniques have been used for studying bacteriorhodopsin under a range of solvent conditions (e.g. folded, SDS, acid, heat-denatured). DEER spectroscopy of SDS-unfolded bacteriorhodopsin has enabled observation of small subpopulations with different intra-helical end-to-end distances and broad distributions of inter-helical distances, suggesting a small degree of heterogeneity in secondary structures but a high degree of heterogeneity in tertiary structures, hence confirming that little native tertiary structure remains in SDS-unfolded bacteriorhodopsin [350]. Kinetic folding experiments of SDS-denatured bacteriorhodopsin in bicelles using pulsed HDX-MS has enabled studying of secondary structure reformation by monitoring the number of protected backbone amides over different folding times ranging from milliseconds to hours, showing that secondary structure formation occurs quite late (on the timescale of > 10 s) on the folding pathway [352]. Oxidative labelling suggests partial unfolding of helices A and D in SDS, and refolded bacterioopsin shows native-like structure but with partial unfolding of helix D, suggesting the importance of retinal in stabilising the structure of this helix [315].

The number and properties of different conformations, the timescales of the sampling motions and the fraction of pSRII population in each conformation can be studied using single-molecule FRET experiments. Millisecond protein-folding dynamics of the small membrane-adhering protein, Mistic, in detergent have been studied using single-molecule FRET. Mistic was labelled with fluorescence donors and acceptors and driven to unfold with urea. A small confocal volume was illuminated, and FRET measurements were performed as single molecules floated in and out of the confocal volume by diffusion. Population peaks at high FRET (folded) and low FRET (unfolded) are observed, with the relative populations changing with different urea concentrations. A significant population of Mistic molecules show a range of intermediate FRET values, suggesting rapid folding/unfolding transitions while passing through the confocal volume. The rates of these folding/unfolding transitions are also measured [353]. An additional advantage of single-molecule approaches is the infinite dilution significantly reduces the potential for aggregation, which is a common problem in membrane protein denaturation studies. Whether single-molecule FRET experiments are able to resolve structural changes in SDS- or acid-denatured pSRII must be

carefully assessed, considering that large amplitude structural changes are unlikely to occur with the protein remaining embedded in micelles under these denaturing conditions.

3.6.3.2 How do retinal binding and Schiff base hydrolysis occur?

Obtaining more detailed structural information on the denatured states and apo sensory opsin could provide clues as to how retinal binding occurs in pSRII. Molecular dynamics simulations suggest a possible binding pathway of retinal through a window between helices E and F of bacteriorhodopsin, through which the β-ionone ring is visible from the outside [354].

Whether Schiff base hydrolysis in SDS- and acid-mediated denaturation of pSRII is facilitated by external or internal water molecule(s) remains unknown. Previous studies have used radioactive ^{18}O incorporation quantified by mass spectrometry to determine that Schiff base hydrolysis of light-activated rhodopsin is mediated by external water molecules [355]. To investigate whether bulk water or small molecules (e.g. hydroxylamine) enter from the extracellular or cytoplasmic side, proteoliposomes are prepared, containing unlabelled hydroxylamine in the liposomal lumen and rhodopsin preferentially oriented with its N-terminus facing the lumen and its C-terminus towards the liposome exterior which is enriched with ^{15}N-labelled hydroxylamine. This experimental set-up shows that small molecules enter rhodopsin exclusively from the extravesicular space (the cytoplasmic side of rhodopsin) [355]. Confirming the mechanism of Schiff base hydrolysis in pSRII would yield further insight onto this important reaction.

It would also be interesting to determine whether retinal binding is reversible for hydroxylamine-bleached sensory opsin, and whether protein dynamics are restored if the protonated Schiff base can be reformed. ^{13}C solid-state NMR on hydroxylamine-bleached bO shows that retinal Schiff base hydrolysis is reversible, and rebinding of the retinal in what was proposed to be trimeric bO leads to restoration of backbone dynamics to yield a comparable spectrum to that of bacteriorhodopsin [343].

Summarising the above, future work building upon the NMR studies in this chapter will require further characterisation of the structures of denatured states and apo sensory opsin, and helix motions in the μs–ms timescale.

4 Assessment of hydrogen bonding through scalar coupling measurements

4.1 Introduction

Hydrogen bonds (H-bonds) play fundamental roles in stabilising protein secondary and tertiary structure, and regulating protein function. Successful detection of transient H-bonds in denatured states and H-bond formation during protein folding would contribute towards our understanding on the structural properties of residual structures and folding/unfolding intermediates.

Several scalar coupling (*J*-coupling) constants report on the presence of H-bonds in proteins. The backbone amide $^{1}J_{NH}$ couplings have typical sizes of about −93 Hz. Upon H-bond formation, $^{1}J_{NH}$ becomes more negative than the amino acid-specific random coil value, whereas residues which are not H-bonded tend to show positive deviations in $^{1}J_{NH}$ values. For ubiquitin, the difference in $^{1}J_{NH}$ ($\Delta^{1}J_{NH}$) against random coil values are small and range from −1.3 Hz to +2.3 Hz [201]. Previous studies have demonstrated residue-specific detection of transient [208] and stable [201] H-bonds in small globular proteins by measuring $^{1}J_{NH}$ using NMR. As *J*-couplings were traditionally only observed in covalently-bonded systems, initial observation of the cross H-bond $^{3h}J_{NC'}$ coupling sparked remarkable interest [207,356]. $^{3h}J_{NC'}$ are small, ranging from −0.1 to −0.9 Hz. $^{3h}J_{NC'}$ can be measured by "long-range" HNCO-COSY (correlation spectroscopy), enabling identification of the H-bond donor, proton and acceptor atoms in one experiment [357]. $^{3h}J_{NC'}$ shows strong dependence on H-bond length [356] and dihedral angles [358], and has been used for monitoring the progressive formation of H-bonds during trifluoroethanol-induced secondary structure formation in an N-terminal segment of Ribonuclease A [359]. One of the drawbacks is that $^{3h}J_{NC'}$ cannot be measured in 3_{10}-helices due to unfavourable geometry [207].

Like any solution-state NMR studies on bio-macromolecules, accurate detection and characterisation of H-bonds require methods which provide high resolution and sensitivity.

However, this is potentially very challenging for membrane protein systems, since detergent-protein complexes can have very large molecular weights of up to ~100−200 kDa. The slow tumbling motion of such large molecules leads to fast decay of the NMR signal and reduction of data quality.

In this chapter, proof-of-concept studies for measuring $^{1}J_{NH}$ and correlating $\Delta^{1}J_{NH}$ with H-bonding patterns were performed in soluble proteins and membrane proteins. $^{1}J_{NH}$ and $\Delta^{1}J_{NH}$ were chosen for their larger magnitudes compared to $^{3h}J_{NC'}$, hence potentially serving as better readouts during the process of optimising different parameters in NMR pulse sequences in order to obtain good resolution and sensitivity. Different experimental methods (and variations thereof) for measuring $^{1}J_{NH}$ were first explored using RalA, a small GTPase with a mixed α/β fold. Detection of H-bonds was then attempted with OmpX, a β-barrel membrane protein, both in its folded state in DPC micelles and in the urea-denatured state, in order to evaluate the accuracy and precision of such measurements in the context of membrane protein folding.

4.2 Theory

Various methods are available for measuring scalar coupling constants. *J*-couplings can be determined directly by measuring resonance splitting (*J*-resolved methods) [360] e.g. the IPAP experiment. In quantitative *J* methods, *J*-couplings are allowed to evolve over a certain amount of time in the pulse sequence, and *J* values are measured by monitoring the modulation of signal intensity originating from a single operator (*J*-modulation) or by obtaining the ratio of the in-phase and antiphase operators after a fixed *J* evolution period (*J*-correlation) [361].

4.2.1 Origins of systematic errors in *J*-coupling measurements

One source of error in *J*-coupling measurements is the differential relaxation rates of in-phase and antiphase magnetisation. The relaxation rate constant of heteronuclear anti-phase magnetisation includes a contribution from longitudinal relaxation of the coupled ^{1}H spin, and is thus greater than the rate constant for in-phase magnetisation [362]. Due to the differential relaxation rates of in-phase and antiphase magnetisation, the observed *J*-coupling (J_{app}) tends to be smaller than the real coupling constant (*J*). This affects *J* measured by resonance splitting and quantitative *J*-correlation (where *J* is extracted from the intensity ratio between in-phase and antiphase operators). The larger the difference between the relaxation rates of the ^{15}N doublet components, the smaller the value of the apparent *J* [363]:

$$\left|J_{\mathrm{app}}\right| = |J| - \frac{(\Delta R)^2}{(2\pi)^2(2|J|)} \tag{4.1}$$

where ΔR is the difference of the relaxation rate constants for antiphase and in-phase transverse magnetisation.

On the other hand, quantitative *J* modulation of a single operator is less sensitive to differential relaxation provided that *J* is large compared to the relaxation rates.

Another main source of systematic error in *J*-coupling measurements is cross-correlated relaxation (CCR). CCR occurs when the random fields from two separate relaxation sources are not independent, hence altering the relaxation rate constants. Chemical shift anisotropy (CSA) and dipolar interactions (DD) are the two dominant relaxation mechanisms (see Chapter 3 for definitions). The time dependence of the local field experienced by the

^{15}N magnetisation derives from thermal motion altering the orientations of the NH bond vector and the ^{15}N CSA tensor relative to the applied magnetic field.

There are two main CCR mechanisms for transverse ^{15}N magnetisation in an amide group:

(1) Chemical shift anisotropy of the ^{15}N with dipolar interaction of the ^{15}N-^{1}H (CSA(^{15}N)/DD$_{NH}$) or nearby ^{1}H which are not attached to ^{15}N (CSA(^{15}N)/DD$_{NR}$)

(2) Dipolar interaction of the ^{15}N-^{1}H with dipolar interaction of the amide ^{15}N and nearby ^{1}H which are not attached to the ^{15}N (DD$_{NH}$/DD$_{NR}$).

In the presence of CSA/DD CCR, the relaxation of the ^{15}N magnetisation depends on the spin state of the ^{1}H magnetisation. The random fields from the DD and CSA interactions are correlated in such a way that they partly cancel each other for the downfield component of the ^{15}N doublet, whereas the two fields reinforce for the upfield component. The downfield component of the ^{15}N doublet relaxes at the difference of autorelaxation (DD + CSA) and CCR whereas the upfield component relaxes at the sum:

Downfield: $$\underbrace{\frac{1}{2}b^2 j(0)}_{\text{DD}} + \underbrace{\frac{2}{45}c_N^2 j(0)}_{\text{CSA}} - \underbrace{\frac{2}{15}c_N b P_2(\cos\theta) j(0)}_{\text{cross-correlation}} \quad (4.2)$$

Upfield: $$\underbrace{\frac{1}{2}b^2 j(0)}_{\text{DD}} + \underbrace{\frac{2}{45}c_N^2 j(0)}_{\text{CSA}} + \underbrace{\frac{2}{15}c_N b P_2(\cos\theta) j(0)}_{\text{cross-correlation}} \quad (4.3)$$

where $j(0)$ is the reduced spectral density function in the slow motion limit, $b = \frac{\mu_0 \gamma_1 \gamma_2 \hbar}{4\pi r^3}$, $c_N = \gamma_N B_0 (\sigma_{N,\parallel} - \sigma_{N,\perp})$ in which $\sigma_{N,\parallel}$ and $\sigma_{N,\perp}$ are the parallel and perpendicular components of the chemical shielding tensor and the subscript N denotes ^{15}N, and $P_2(\cos\theta) = \frac{1}{2}(3\cos^2\theta - 1)$ in which θ is the angle between the principal axis of the chemical shielding tensor and the N-H vector.

CSA(^{15}N)/DD$_{NH}$ CCR results in an asymmetric ^{15}N doublet, in which the upfield resonance is broader and shorter (i.e. the anti-TROSY line) while the downfield resonance is sharper and taller (i.e. the TROSY line). The asymmetry increases as the correlation time becomes greater. This phenomenon can be used advantageously in transverse relaxation optimised spectroscopy (TROSY) [364], in which selection of the narrow line in one or both dimensions of the doublet results in spectral resolution enhancement. *J*-couplings derived

from measuring resonance splitting are affected by inaccuracies in determining the frequency values of broad anti-TROSY lines [365]. CSA(^{15}N)/DD$_{NR}$ CCR arises in the presence of dipolar interaction between ^{15}N and a remote proton, H^R, leading to observation of a doublet of doublets. Given the small magnitude of ^{15}N-^{1}H^R dipolar interactions, the effects of CSA(^{15}N)/DD$_{NR}$ CCR usually manifests as slight skewing of the peaks in the ^{15}N doublet.

While it is assumed in the above discussion that only the ^{15}N has a significant CSA, in practice, the ^{1}H also has a CSA. Although the magnitude of the ^{1}H CSA is about one-tenth of that for the ^{15}N CSA, the larger ^{1}H gyromagnetic ratio entails that CSA(^{1}H)/DD contributes to a similar extent as CSA(^{15}N)/DD [199]. However, since ^{1}H is only present as longitudinal magnetisation (H_z), CSA(^{1}H)/DD is effectively negligible in the slow tumbling limit.

The imaginary component of the spectral density function associated to CSA/DD CCR also induces a change in the *J*-coupling, known as dynamic frequency shift (DFS). DFS arises from a differential shift of the chemical shifts of the two spin states [366]. DFS decreases with increasing rotational correlation time and greater magnetic field strength, and contributes additively to the apparent *J*-coupling [367] measured by both the resonance splitting and the quantitative *J* methods [368]:

$$\mathrm{DFS} \approx \left(\frac{1}{10\pi}\right)\left(\frac{\mu_0 \hbar \gamma_N \gamma_H}{4\pi r_{NH}^3}\right)\left(\sigma_{N,\parallel} - \sigma_{N,\perp}\right)(3\cos^2\theta - 1) \times \frac{1}{\left(1+\frac{1}{(\omega_N \tau_c)^2}\right)} \qquad (4.4)$$

Unlike CSA/DD CCR, which can be suppressed by judicious choice of 180° pulses, DFS cannot be suppressed [366]. In view of the high magnetic field strengths utilised and the rotational correlation times (~20 ns) of the proteins studied in this chapter, DFS is expected to be constant at about –0.5 Hz [369].

DD$_{NH}$/DD$_{NR}$ CCR refers to relaxation interference between the amide ^{15}N-^{1}H dipolar interaction and dipolar interaction between ^{15}N and a remote proton, H^R, which is not directly bonded to ^{15}N [365,370]. In order for DD$_{NH}$/DD$_{NR}$ CCR to affect $^1J_{NH}$ measurements, H^R must be *J*-coupled to ^{15}N ($J_{N\text{-}HR}$). For protein backbone amides, the most plausible candidates for H^R are the side-chain H$_\alpha$ and H$_\beta$ atoms. As a result of $J_{N\text{-}HR}$, the ^{15}N multiplet for the three-spin system (H^N-N$\cdots$H^R) is a doublet of doublets (Figure 4.1). The impact of

DD_{NH}/DD_{NR} CCR on *J*-coupling measurements depends on the size of $J_{N\text{-}HR}$ and geometry of the H^N-N···H^R three-spin system. The dipolar contributions to the relaxation rates of the ^{15}N doublet components are proportional to:

$$\frac{1}{r_{NH}^6} \pm \frac{2P_2 \cos\theta}{r_{NH}^3 r_{N-HR}^3} + \frac{1}{r_{N-HR}^6} \quad (4.5)$$

where r_{NH} is the N-H bond length, $r_{N\text{-}HR}$ is the N-H^R internuclear distance, and θ is the angle between the ^{15}N-$^1H^N$ and ^{15}N-$^1H^R$ vectors.

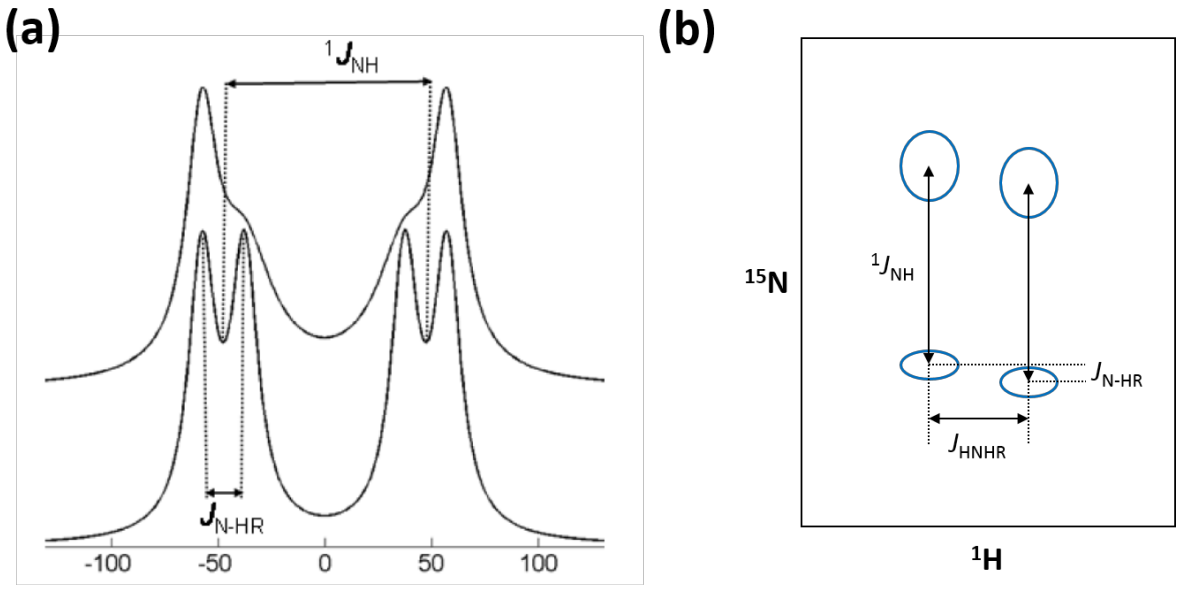


Figure 4.1: The impact of dipole/dipole cross-correlated relaxation on $^1J_{NH}$ measurements
(a) In the presence of a remote proton, H^R, which is *J*-coupled to but not attached to the amide ^{15}N, dipole/dipole cross-correlated relaxation between $^1H^R$-^{15}N and ^1H-^{15}N gives rise to asymmetric peaks (top spectrum). For visual purposes, the outer and inner linewidths were set as 17 and 33 Hz, respectively, $J_{N\text{-}HR}$ was set as 20 Hz, and relaxation interference was exaggerated. The lower spectrum was simulated in the absence of relaxation interference and with linewidth of 15 Hz. Retrieved from [370]. **(b)** Schematic diagram showing the E.COSY-type fine structure in ^1H-^{15}N HSQC correlations due to the presence of both J_{HNHR} and $J_{N\text{-}HR}$. Adapted from [370].

Depending on the backbone ϕ and side-chain χ_1 torsion angles, intraresidue H_β can be proximate to ^{15}N and $^3J_{NH\beta}$ can be as large as 4 Hz, resulting in a dominant effect of N-H^N/N-H_β DD/DD CCR on $^1J_{NH}$ measurements. In comparison, the effects of N-H^N/N-H_α are small, because intraresidue $^2J_{NH\alpha}$ and interresidue $^3J_{NH\alpha}$ are small and interresidue N-H_α are > 3Å in helices.

H^R can also be *J*-coupled to the amide ^1H (J_{HNHR}). Under isotropic conditions, only the intraresidue $^3J_{HNH\alpha}$ might be sufficiently large to be resolved. Combined with $J_{N\text{-}HR}$, this gives rise to an E.COSY-type fine structure (Figure 4.1b). $J_{N\text{-}HR}$, J_{HNHR} and the combined fine structure are often unresolved. In principle, both ^{15}N doublet components receive equal contributions. However, in practice, the difference in the linewidths of doublet components due to CSA(^{15}N)/DD_{NH}, combined with the unresolved fine structure due to DD_{NH}/DD_{NR}, leads to asymmetry of the lineshape in the ^{15}N dimension. The skewed peak shapes could give rise to problems with peak fitting, thereby adversely impacting the accuracy of $^1J_{NH}$ measurements.

To summarise, for isotropic samples, errors due to DD_{NH}/DD_{NR} depend on $^2J_{NH\alpha}$, $^3J_{NH\beta}$ and $^3J_{HNH\alpha}$ and internuclear distances and vector angles involved in $^{15}N_{(i)}$-$^1H_{\alpha(i-1)}$, $^{15}N_{(i)}$-$^1H_{\alpha(i)}$ and $^{15}N_{(i)}$-$^1H_{\beta(i)}$ dipole interactions.

In addition, magnetisation can be transferred between operators by CCR. Coherence transfer is typically mediated by *J*-coupling evolution. For a spin-system with two spin ½ nuclei, coherence transfer via *J*-coupling results in the generation of single quantum in-phase and antiphase magnetisation, the ratio of which depends on the duration of evolution. Coherence transfer can also occur via relaxation, whereby "forbidden" multiple-quantum and zero-quantum coherences are generated, giving rise to *J*-independent coherence transfer. Some of these "forbidden" terms can be subsequently transferred back to yield observable terms. For example, N-H^N/N-H_β DD/DD CCR in the presence of $^3J_{NH\beta}$ can cause the conversion of terms antiphase with respect to the $^1H_\beta$ (*K*) spin back into observable magnetisation [371,372]:

$$-4N_xH_zK_z \xrightarrow{DD/DD} -N_x \tag{4.6}$$

$$-2N_yK_z \xrightarrow{DD/DD} -2N_yH_z \tag{4.7}$$

This mechanism could lead to undesirable $^1J_{NH}$-independent contribution towards signal intensities, and under-/over-estimation of the measured *J* value in Quantitative *J* experiments (see Section 4.2.3 for product operators).

The simplest method to avoid contamination of one-bond coupling measurements by relaxation interference effects would be to utilise perdeuterated sample with back-exchanged amide protons, or to suppress CCR using the appropriate selective pulse sequence.

4.2.2 ^{15}N IPAP-HSQC

The original ^{1}H-^{15}N IPAP-HSQC experiment records two [^{1}H,^{15}N] HSQC spectra in an interleaved manner without ^{1}H decoupling in the ^{15}N dimension [373]. The first spectrum, known as in-phase (IP), is a regular HSQC with ^{1}H coupled to ^{15}N, and the two components of each ^{15}N doublet have the same sign. The second spectrum, known as antiphase (AP), is generated by inserting a ^{15}N refocusing period $\Delta \approx (2J_{\mathrm{NH}})^{-1}$ before the t_1 evolution period, and the ^{15}N doublet components have opposite signs. A suitable scaling factor is applied to the AP spectrum to account for differences in peak intensities. Sub-spectra containing either only the upfield or the downfield ^{15}N doublet component is generated by adding or subtracting the IP and AP spectra.

While perdeuteration is the easiest method to achieve high accuracy in *J*-coupling measurements, Yao *et al*. have demonstrated that $^{1}J_{\mathrm{NH}}$ measurements of comparable accuracy can be achieved in protonated samples. Several modifications have been introduced in the implementation reported by Yao *et al*. [370]:

1) utilise pulse sequences of similar lengths for recording the IP and AP spectra to reduce differences in peak intensities (see yellow box in Figure 4.2);
2) band-selective decoupling of aliphatic protons to refocus $J_{\mathrm{N–HR}}$, thereby suppressing DD/DD CCR; and
3) gradient-enhanced coherence transfer for better spectral quality.

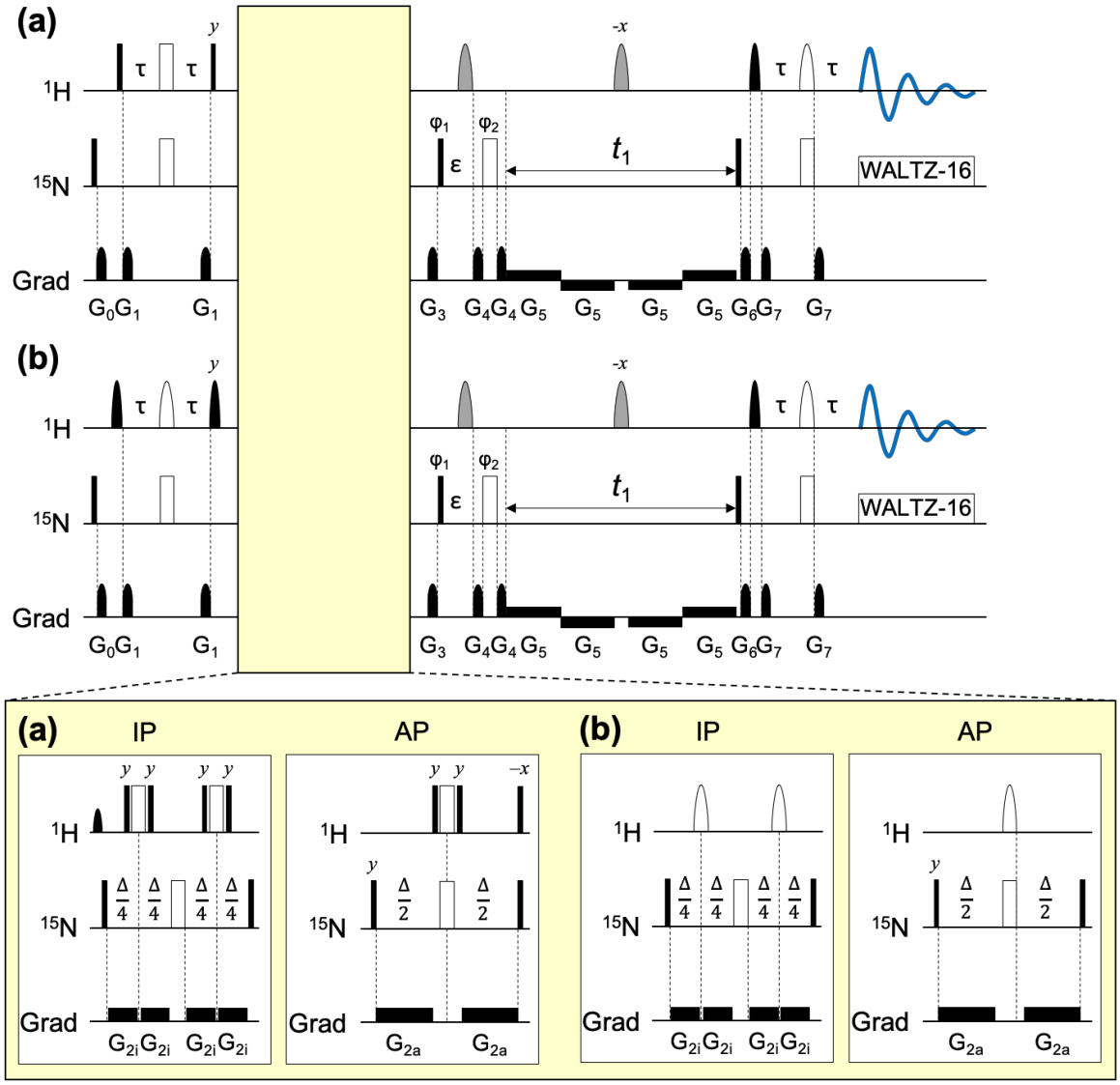


Figure 4.2: Pulse sequences for ^{15}N-IPAP-HSQC (a) and ^{15}N-SOFAST-IPAP-HSQC (b)
Narrow black pulses and wide white pulses correspond to 90° and 180° flip angles, respectively. Pulses for which the RF phase is not indicated are applied along the *x* axis. Pulses in the yellow box are executed for the in-phase (IP) and antiphase (AP) spectra of ^{15}N-IPAP-HSQC (a) and ^{15}N-SOFAST-IPAP-HSQC (b). The grey shaped ^{1}H pulses are IBURP inversion pulses with duration of 1.62 ms, and serve to decouple H^{α} from ^{15}N. The black and white shaped ^{1}H pulses in the reverse INEPT are EBURP 90° excitation pulses with duration of 2.304 ms and 180° Q3 gaussian cascade pulses with duration of 1.76 ms. Delay durations: τ = 2.69 ms (a) or 2.75 ms (b); Δ = 5.38 ms; ε = 0.227 ms. Pulsed field gradients $G_{0,1,3,4,6,7}$ are smooth square shaped with 39%, 11%, 25%, 27%, 29%, 29% strength, and $G_{2i,2a,5}$ are rectangular with 7%, 7%, 2.1% strength, where 100% is 60 G/cm. Gradient durations: $G_{0,1,2i,2a,3,4,5,6,7}$ = 1.5, 0.7 (a) or 0.3 (b), 1.345, 2.69, 1.5, 1, $t_1/4$, 0.5, 2.5 ms. Delay and gradient durations are quoted for ^{1}H frequency of 500 MHz. The phase cycling is: $\phi_1 = x, -x$; $\phi_2 = 2(x), 2(y), 2(-x), 2(-y)$; receiver = $2(x, -x, -x, x)$. Quadrature detection in the t_1 dimension is achieved by incrementing ϕ_1. **(a)** The shaped water flip-back pulse in IP is g3.800 with duration of 2.1 ms. The bold hatched bars are 90_y–210_x–90_y composite 180° pulses. **(b)** The black and white shaped ^{1}H pulses in the INEPT element are amide-selective PC9 90° excitation pulses with duration of 3 ms and 180° Q3 gaussian cascade pulses with duration of 1.76 ms, respectively.

The initial INEPT magnetisation transfer followed by 90° pulses on ^{1}H and ^{15}N yields antiphase transverse ^{15}N magnetisation: $2H_zN_y$ for the IP pulse sequence and $-2H_zN_x$ for the AP pulse sequence. During the Δ period, the magnetisation remains as $2H_zN_y$ in the IP pulse sequence because both *J*-coupling and chemical shift evolution are refocused, whereas $-2H_zN_x$ evolves into $-N_y$ in the AP pulse sequence:

$$-2H_zN_x \longrightarrow -2H_zN_x\cos(\pi J\Delta) - \boldsymbol{N_y\sin(\pi J\Delta)} \tag{4.8}$$

where $\Delta \approx (2J_{\mathrm{NH}})^{-1}$.

After the 90° pulses on ^{15}N, both the ^{15}N chemical shift (Ω_N) and $^{1}J_{\mathrm{NH}}$ are evolved during the t_1 evolution:

AP:
$$N_y \xrightarrow{t_1} \cos(\pi Jt_1)\left[N_y\cos(\Omega_N t_1) - N_x\sin(\Omega_N t_1)\right]$$
$$-\sin(\pi Jt_1)\left[2H_zN_x\cos(\Omega_N t_1) + \boldsymbol{2H_zN_y\sin(\Omega_N t_1)}\right] \tag{4.9}$$

IP:
$$-2H_zN_y \xrightarrow{t_1} \cos(\pi Jt_1)\left[\boldsymbol{-2H_zN_y\cos(\Omega_N t_1)} + 2H_zN_x\sin(\Omega_N t_1)\right]$$
$$+\sin(\pi Jt_1)\left[N_x\cos(\Omega_N t_1) + N_y\sin(\Omega_N t_1)\right] \tag{4.10}$$

where only the terms shown in bold result in the desired signals.

Band-selective ^{1}H decoupling of aliphatic protons during the t_1 evolution period removes the net evolution of $J_{\text{N-HR}}$ couplings, thereby removing the unresolved E.COSY-type fine structure and associated lineshape asymmetry in the ^{15}N dimension resulting from protons with non-zero $^{3}J_{\mathrm{HNHR}}$ to $^{1}\mathrm{H}^{\mathrm{N}}$ and $^{n}J_{\mathrm{NH}}$ to ^{15}N. In principle, DD/DD CCR effects could be removed by the application of one band-selective ^{1}H inversion pulse at the mid-point of the t_1 evolution. However, in practice, applying a selective pulse on aliphatic proton spins leads to off-resonance effects on the amide proton spins, leading to a decrease in the apparent magnitude of $^{1}J_{\mathrm{NH}}$. To first order approximation, the off-resonance effects can be refocused with an additional selective 180° pulse. Hence, two selective 180° pulses were applied to decouple $J_{\text{N-HR}}$, and ^{15}N evolves with opposite coherence orders during each of these pulses.

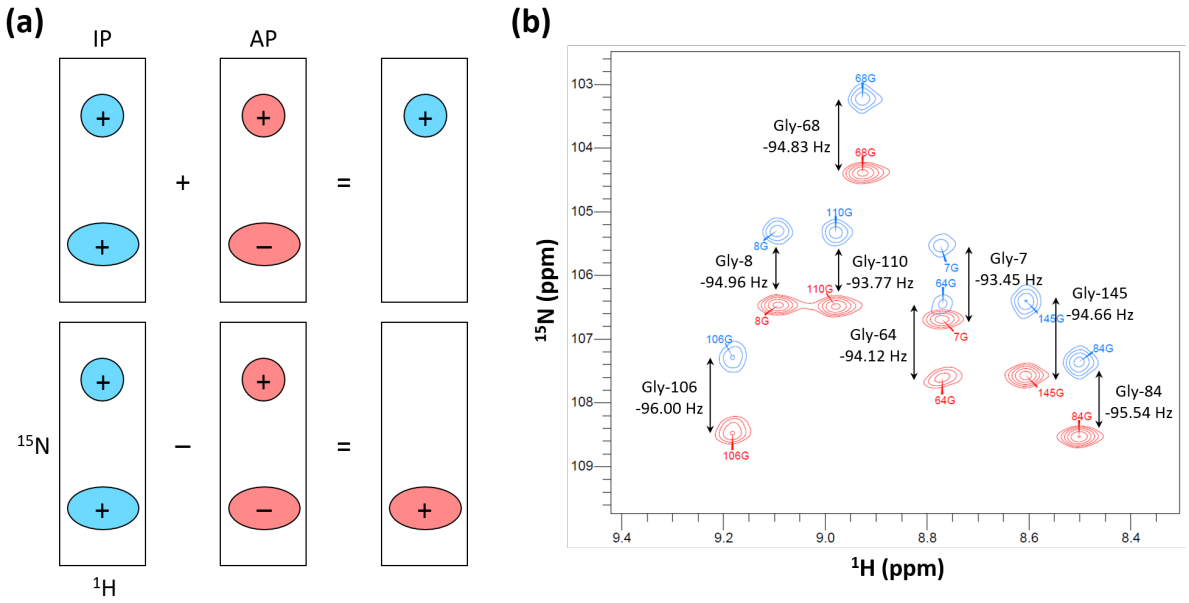


Figure 4.3: IPAP
(a) Schematic diagram of the generation of IPAP sub-spectra by the addition of the in-phase (IP) and anti-phase (AP) data sets. **(b)** Selected region of an IPAP-HSQC spectrum of OmpX. The red and blue sub-spectra of the ^{15}N doublets were generated through addition and subtraction, respectively, of the recorded in-phase (IP) and anti-phase (AP) data sets.

Magnetisation contributing towards the desired signals (see Equations 4.9–4.10) are transferred back to ^{1}H for detection via a reverse INEPT. The observable signals are:

AP: $$H_x \sin(\Omega_N t_1) \sin(\pi J t_1) = \frac{1}{2} H_x \{\cos[(\Omega_N - \pi J)t_1] - \cos[(\Omega_N + \pi J)t_1]\} \quad (4.11)$$

IP: $$H_x \cos(\Omega_N t_1) \cos(\pi J t_1) = \frac{1}{2} H_x \{\cos[(\Omega_N - \pi J)t_1] + \cos[(\Omega_N + \pi J)t_1]\} \quad (4.12)$$

Addition of the IP and AP spectra yields the subspectrum with only the upfield ^{15}N doublet component with ^{15}N chemical shift of $(\Omega_N - \pi J)$; subtraction of the IP and AP spectra yields the subspectrum with only the downfield component at $(\Omega_N + \pi J)$ (Figure 4.3).

Quadrature detection is achieved by simultaneously incrementing the phases for ϕ_1.

Various modifications were introduced by Dr. Daniel Nietlispach to further improve the sensitivity. Amide-selective shaped pulses were used during magnetisation transfer in SOFAST-IPAP-HSQC. Combined with the pulsed field gradients, this entails that the spin states of protons which are not attached to ^{15}N do not produce detectable signals.

The product operators for the desired magnetisation transfer pathway are the same for SOFAST-IPAP-HSQC and IPAP-HSQC.

Optimal sensitivity also relies on active restoration of the water signal to the *z*-axis during execution of the pulse sequence. For IPAP-HSQC, this is achieved using water flip back pulses at the end of the *J*-coupling evolution period as detailed by Yao *et al.* [370]: a water-selective 90° sine-bell-shaped pulse is applied to restore water magnetisation to +*z* in the IP experiment, whereas a non-selective ^{1}H 90°(–*x*) purge pulse is applied in the AP experiment. The utilisation of amide-selective pulses in SOFAST-IPAP-HSQC indicates that the water signal is not excited. Hence, water flip back pulses were removed by Dr. Daniel Nietlispach.

4.2.3 Quantitative *J* modulation

Quantitative *J* methods can be classified as *J*-correlation or *J*-modulation methods. *J* is determined from *J*-correlation experiments by obtaining the ratio of the in-phase and antiphase operators after a fixed *J* evolution period, or by comparing peak intensities against a separately acquired reference spectrum. The constant *J* evolution period would ensure that the length of the pulse sequence and signal intensities are consistent across all experiments. However, large errors could result from difficulties in accurately determining low signal intensities derived from magnetisation which are antiphase at the end of the *J* evolution period. In *J*-modulation experiments, *J* values are measured by monitoring modulation of the signal intensity originating from a single operator (*J*-modulation). An advantage is that only signal derived from in-phase magnetisation after the *J* evolution period is required, hence broad linewidths associated with antiphase operators can be circumvented. Moreover, errors in the derived *J*-couplings can be easily estimated based on the quality of the multi-point fitting of intensities. However, multiple experiments with different *J* evolution times are required to obtain signal intensity modulation, indicating that error would primarily originate from sensitivity loss from transverse relaxation over prolonged *J* evolution times.

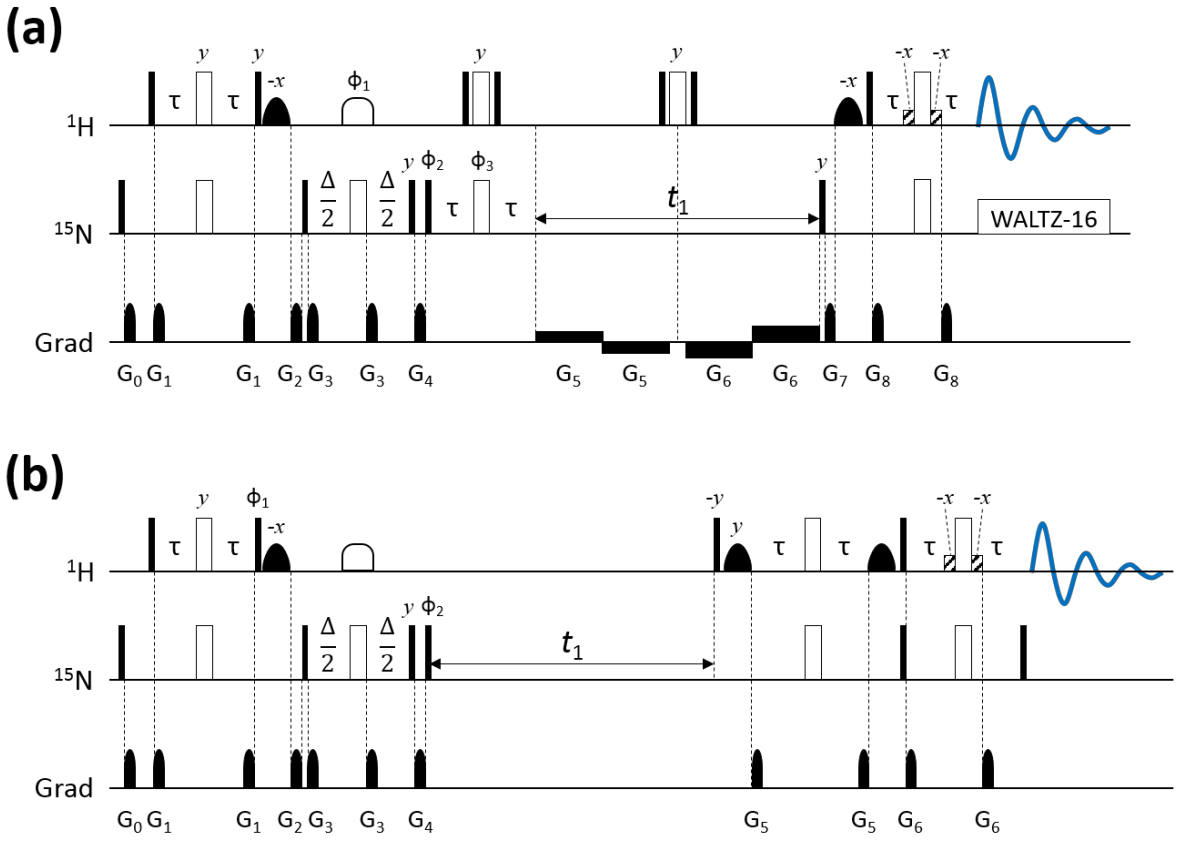


Figure 4.4: Pulse sequence for $^1J_{NH}$-modulated HSQC (a) and $^1J_{NH}$-modulated TROSY-HSQC (b) Narrow black pulses and wide white pulses correspond to 90° and 180° flip angles, respectively. Pulses for which the RF phase is not indicated are applied along the *x* axis. The black shaped ^{1}H pulses are ESNOB water flip back pulses with duration of 2.1 ms. The white shaped ^{1}H pulse during the *J* evolution period Δ is a REBURP pulse with duration of 1.92 ms. The low power ^{1}H 90° pulses are part of the WATERGATE water suppression scheme, and are applied for 1.4 ms each. Delay durations: τ = 2.65 ms (a) or 2.75 ms (b). 15 data points within the ranges 11.60–37.20 ms, 22.24–47.84 ms, or 32.88–58.48 ms were taken for Δ period. **(a)** The bold hatched bars are 90_x–210_y–90_x composite 180° pulses. Pulsed field gradients $G_{0,1,3,4,6,7,8}$ are smooth square shaped with 37%, 17%, 23%, 25%, 35%, 39%, 27% strength, and $G_{5,6}$ are rectangular with 1.1%, 2.1% strength, where 100% is 60 G/cm. Gradient durations: $G_{0,1,2,3,4,5,6,7,8}$ = 1.5, 0.7, 1.3, 1.5, 1.9, t_1/4, t_1/4, 2.5, 0.6 ms. The phase cycling is: ϕ_1 = 2(*x*), 2(*y*), 2(–*x*), 2(–*y*); ϕ_2 = *x*, –*x*; ϕ_3 = 2(*x*), 2(*y*), 2(–*x*), 2(–*y*); receiver = 2(*x*, –*x*, –*x*, *x*). Quadrature detection in the t_1 dimension is achieved by simultaneously incrementing the phases ϕ_2 and ϕ_3. **(b)** Pulsed field gradients $G_{0,1,2,3,4,5,6}$ are smooth square shaped with 39%, 13%, 37%, 27%, 39%, 27%, 39% strength, where 100% is 60 G/cm. Gradient durations: $G_{0,1,2,3,4,5,6}$ = 1.3, 0.7, 2.5, 1.3, 2.5, 0.5, 0.6 ms. The phase cycling is: ϕ_1 = 4(–*y*), 4(*y*); ϕ_2 = *y*, –*y*, –*x*, *x*, *y*, –*y*, *x*, –*x*; receiver = *x*, –*x*, *y*, –*y*, –*x*, *x*, –*y*, *y*. Quadrature detection in the t_1 dimension is achieved by incrementing ϕ_2.

Tjandra *et al.* used a pulse sequence in which the peak intensity is modulated as an exponentially decaying cosine function with respect to the $^1J_{NH}$ couplings and the T_2 relaxation rate of ^{15}N nuclei [369]. Slight modifications to the pulse sequence were made by Dr. Daniel Nietlispach such that the peak intensity is modulated as an exponentially decaying

sine function (Figure 4.4). Rather than detecting signal generated from antiphase magnetisation at the end of the variable *J* evolution period (Δ in Figure 4.4), the modified pulse sequence allows selective detection of signal originating from magnetisation which is in-phase after Δ. It was hypothesised that detecting only the signal that originates from this refocused magnetisation will minimise the effects of pulse imperfections.

The INEPT element and 90° pulses on ^{1}H and ^{15}N enable magnetisation transfer from ^{1}H to antiphase ^{15}N magnetisation. The resulting operator $2H_zN_y$ undergoes *J* coupling evolution during Δ:

$$2H_zN_y \longrightarrow 2H_zN_y\cos(\pi J\Delta) - \boldsymbol{N_x\sin(\pi J\Delta)} \tag{4.13}$$

The REBURP pulse in the mid-point of Δ selectively inverts $^{1}H^{N}$, hence eliminating J_{N-HR} coupling and the associated DD/DD CCR effects.

Only the $-N_x\sin(\pi J\Delta)$ term is carried through the rest of the pulse sequence. After two ^{15}N 90° pulses and an INEPT element to generate antiphase magnetisation, the ^{15}N chemical shift is then evolved during t_1:

$$2H_zN_x\sin(\pi J\Delta) \longrightarrow \sin(\pi J\Delta)\left[\boldsymbol{2H_zN_x\cos(\Omega_N t_1)} + N_y\sin(\Omega_N t_1)\right] \tag{4.14}$$

The $2H_zN_x\cos(\Omega_N t_1)\sin(\pi J\Delta)$ term is transferred back to ^{1}H via the reverse INEPT element and the observable signal is:

$$H_x\cos(\Omega_N t_1)\sin(\pi J\Delta) \tag{4.15}$$

Quadrature detection in the t_1 dimension is achieved by simultaneously incrementing the phases of the ^{15}N 90° and 180° pulses in the τ period before t_1 evolution.

In the presence of pulse imperfections, DD/DD CCR would arise due to incomplete refocusing of J_{N-HR}. The magnetisation would evolve during the *J*-coupling evolution period as follows:

$$2H_z^N N_y \cos(\pi J_N \Delta) \cos(\pi J_R \Delta) - N_x \sin(\pi J_N \Delta) \cos(\pi J_R \Delta)$$
$$-2H_z^R N_y \sin(\pi J_N \Delta) \sin(\pi J_R \Delta) - 4H_z^N H_z^R N_x \cos(\pi J_N \Delta) \sin(\pi J_R \Delta) \quad (4.16)$$

where J_N is the amide $^1J_{NH}$ and J_R is the *J*-coupling between amide ^{15}N and H_α or H_β.

As a result, the collected operator is modulated as a function of:

$$-N_x[\sin(\pi J_N \Delta) \cos(\pi J_R \Delta) \cosh(\Gamma \Delta) + \cos(\pi J_N \Delta) \sin(\pi J_R \Delta) \sinh(\Gamma \Delta)] \quad (4.17)$$

where Γ is the DD/DD CCR rate between the amide ^{15}N-^1H dipole and the ^{15}N-$H_{\alpha/\beta}$ dipole.

Water suppression is achieved by Tjandra *et al*. by a single WATERGATE water suppression scheme incorporated in the reverse INEPT. Additional ESNOB water flip back pulses were applied after the initial INEPT and before the reverse INEPT to improve water suppression.

While the pulse sequence has been discussed above with references to the $^1J_{NH}$-modulated HSQC implementation, these general principles would also apply to the TROSY implementation, which takes advantage of the TROSY effect due to CSA/DD CCR to obtain narrow linewidths for accurate intensity determination. The salient features of the pulse sequence remain similar except for utilisation of the TROSY effect during t_1 evolution and acquisition (t_2). The lack of refocusing ^1H pulse during t_1 evolution in $^1J_{NH}$-modulated TROSY-HSQC enables simultaneous chemical shift and coupling evolution under the influence of CSA/DD CCR. Only the sharp ^{15}N line in the t_1 dimension is maintained and correlated with the sharp ^1H line in the t_2 dimension. This is achieved through the single-transition to single-transition polarisation transfer element (ST2PT) [364], which transfers ^{15}N coherence back to ^1H before acquisition.

15 experiments were acquired in which the Δ period for *J* evolution is varied. The resulting exponentially-decaying sinusoidal modulation of signal intensity (*I*) goes through three maxima/troughs and two zero-crossings as a function of Δ:

$$I = A\sin(\pi J\Delta)\, e^{-C\Delta} \tag{4.18}$$

where *J* is the $^{1}J_{NH}$ coupling constant, C is the decay rate (R_2) of the ^{15}N magnetisation as a result of transverse relaxation, and A is a signal amplitude scaling factor.

Figure 4.5 shows the fitted sinusoidal intensity modulation for three residues from OmpX in DPC and a data table listing the fitted parameters, *J*, A and C. Residue Tyr-95 shows the best fit, whilst Arg-133 represents an average fit compared to other residues in the data set. Residues such as Tyr-30 are omitted from further analysis due to insufficient number of data points and inadequate sampling of only one zero-crossing in the intensity modulation curve. The goodness-of-fit is expressed in terms of the 2.5% and 97.5% confidence intervals of *J* (denoted as $J_{2.5\%}$ and $J_{97.5\%}$, respectively, in Table 4.1), where a small difference between $J_{2.5\%}$ and $J_{97.5\%}$ indicates a good fit of the data with the sinusoidal decay function. Table 4.1 shows the output values for *J*, $J_{2.5\%}$, $J_{97.5\%}$, A and C from the data fitting.

	J	$J_{2.5\%}$	$J_{97.5\%}$	**A**	**C**
Arg-133	–93.657	–92.250	–95.261	2079407	50.057
Tyr-30	–95.522	–88.035	–106.888	967689	42.757
Tyr-95	–92.389	–92.220	–92.562	7432110	30.518

Table 4.1: Examples of fitted $^{1}J_{NH}$ parameters
Output values for *J*, A and C from the data fitting of $^{1}J_{NH}$-modulated signal intensities of Arg-133, Tyr-30 and Tyr-95. $J_{2.5\%}$ and $J_{97.5\%}$ denote the 2.5% and 97.5% confidence intervals for the fitted *J* value, respectively.

Quantitative *J* modulation is theoretically less prone to error than frequency-based measurements, since CSA/DD CCR affects accurate determination of peak positions but not signal intensity modulation. This makes quantitative *J* modulation methods preferable for the measurement of small homo- and heteronuclear couplings. Sensitivity loss from transverse relaxation and the effects of transfer of "forbidden" terms adversely affect the accuracy of *J* values determined by both IPAP and quantitative *J* modulation methods.

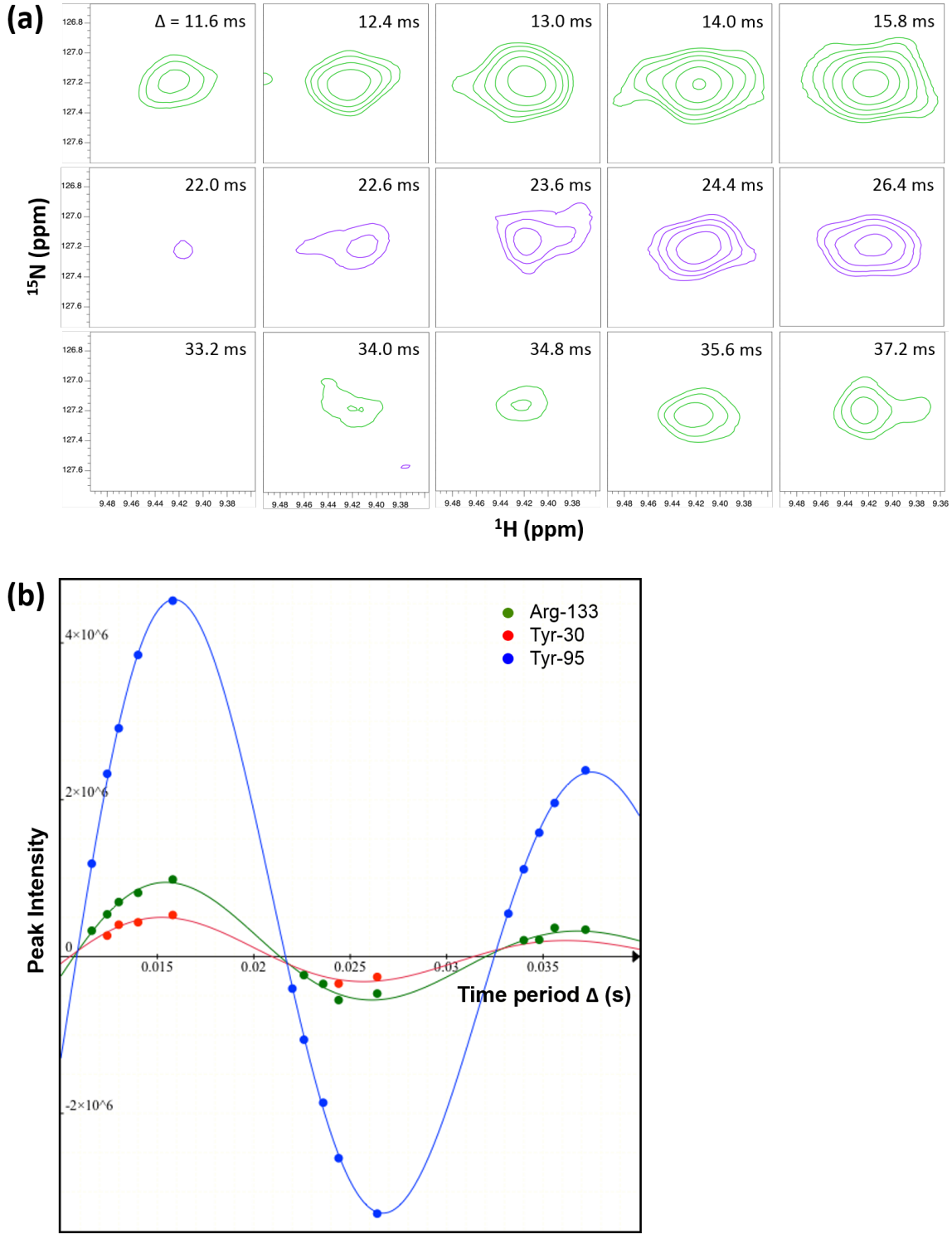


Figure 4.5: $^{1}J_{NH}$-modulated HSQC
(a) Small sections of the 2D [^{1}H-^{15}N] HSQC spectra of OmpX showing the peak for Arg-133 recorded at different *J* evolution periods (Δ) ranging 11.6–37.2 ms. **(b)** Modulation of peak intensity as a function of Δ. Measured peak intensities are shown as dots, and are overlaid with the fitted sinusoidal decay function.

4.3 Results

4.3.1 Proof-of-concept & optimisation: $^1J_{NH}$ of RalAΔC-GDP

GDP-bound C-terminally truncated RalA (RalAΔC), consisting of residues 1–184, was chosen for proof-of-concept studies on $^1J_{NH}$ measurements in proteins with molecular weight of >10 kDa. The sample was made by Dr. Arooj Shafiq [374]. RalAΔC contains five α-helices and six β-sheets, and has a molecular weight of 20.9 kDa. As H-bonds in α-helices and β-sheets have different geometries and bonding strengths, the mixed α/β structure of RalAΔC enables the accuracy and precision of $^1J_{NH}$ values to be assessed for different secondary structures.

Backbone-to-backbone H-bonds were predicted from the crystal structure of RalA (Ingrid Vetter, Max Planck Institute of Molecular Physiology) using the "Optimal Hydrogen Bonding Network" option of the WhatIf server [202] (Figure 4.6). WhatIf locates the positions of polar hydrogen atoms in protein structures by optimising the total H-bond energy by means of an empirical H-bond force field. This method takes into account one-to-one and bifurcated H-bonds. WhatIf accounts for potential crystallographic mis-assignments of His, Gln and Asn side-chain conformations by allowing 180° change in the last χ angle of these residues during H-bond energy optimisation.

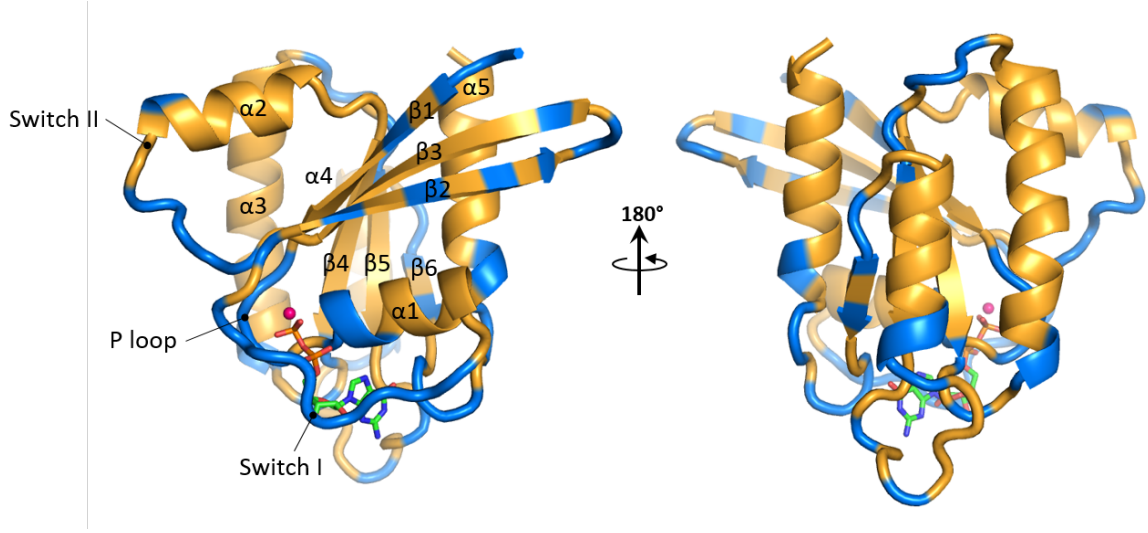


Figure 4.6: H-bonds predicted by the WhatIf server based on the crystal structure of RalAΔC
Backbone-to-backbone H-bonded residues are coloured yellow, and residues which are not involved in backbone-to-backbone H-bonds are coloured in blue. Crystal structure was provided by Ingrid Vetter (Max Planck Institute of Molecular Physiology).

Different implementations of IPAP-HSQC were recorded (Table 4.2):

IPAP-HSQC implementation		No. (%) of residues with reliable $^1J_{NH}$ measurements
IPAP-HSQC	32 scans [A]	121 (72%)
	32 scans [B]	122 (72%)
	112 scans	119 (70%)
SOFAST-IPAP-HSQC	64 scans [A]	118 (70%)
	64 scans [B]	118 (70%)

Table 4.2: Residues with reliable $^1J_{NH}$ measurements in different IPAP-HSQC implementations
[A] and [B] denote separate repeats of the same implementation for empirical assessment of the error. All experiments were recorded at 298 K on a 500 MHz spectrometer.

90° square excitation pulses are used in IPAP-HSQC, whereas the SOFAST implementation utilises 90° shaped excitation pulses. Different number of scans were used to determine the effects of increased signal-to-noise ratio (SNR) on error in $^1J_{NH}$ measurements.

Since resonance splitting methods rely heavily on accurate determination of peak positions, overlapping resonances for which peak positions cannot be obtained by automatic parabolic fitting by CCPN Analysis [375] were excluded from further analysis. Applying this criterion, reliable $^1J_{NH}$ values were obtained for ~120 out of the 169 assignable backbone amide resonances [374] across different IPAP-HSQC implementations (Table 4.2), providing acceptable coverage (~70%) across the protein.

$^1J_{NH}$ for RalAΔC-GDP ranged from –90 to –96 Hz (Figure 4.7a), similar to the range of −91.1 to −95.6 Hz obtained from previous studies on ubiquitin [201]. Δ^1J_{NH} values spanned between ± 2.9 Hz. The experimental precision of IPAP-HSQC experiments was assessed by calculating the sum of the combined $\{^1H,^{15}N\}$ chemical shift differences for both components of the ^{15}N doublet between two experimental repeats (Figure 4.7b). The chemical shift difference for each component of the ^{15}N doublet is calculated as outlined in Chapter 3. Most residues showed a total chemical shift difference (summed across the ^{15}N doublet) of 2–3 Hz

between experimental repeats. These errors are very big, considering that Δ^1J_{NH}, the parameter of interest, is typically between –0.5 Hz and +0.5 Hz [201].

To ascertain the origin of large errors, the sum of the root mean squared deviation (RMSD) in peak positions for both components of the ^{15}N doublet was calculated (Figure 4.7c). Peak position RMSD for each component of the ^{15}N doublet was calculated using signal-to-noise ratios and transverse relaxation rates of individual residues:

$$\mathrm{RMSD} = \frac{1}{\mathrm{SNR}}\left(\frac{0.26}{t_{1max}} + 0.067C\right) \tag{4.19}$$

where SNR is the signal-to-noise ratio of a peak, $t_{1\mathrm{max}}$ is the maximum increment of the t_1 evolution period in seconds, and C is the transverse relaxation rate constant of the ^{15}N magnetisation estimated from $^1J_{NH}$-modulated HSQC.

The relationship was derived by Dr. Daniel Nietlispach based on simulated resonances reported by Kontaxis *et al* [376]. 2D resonances were simulated with Gaussian noise added in the time domain to yield consistent SNR = 20. The accuracy for determining peak position increases linearly with SNR and with R_2. It is important to use relaxation rates derived from spin-echo type experiments rather than CPMG, because relaxation in spin echo-type experiments occurs at the average relaxation rate of in-phase and antiphase operators, whereas relaxation occurs at the rate of in-phase operator in CPMG experiment. This is clearly illustrated by the lower R_2 derived from CPMG [374] compared to the decay rate (C) obtained by fitting peak intensities from $^1J_{NH}$-modulated HSQC peak using exponentially-decaying sinusoidal function (Figure 4.7d).

For IPAP-HSQC (32 scans), most residues yielded RMSD of ~ 0.1 Hz, although a few select residues showed higher RMSD of up to ~ 0.5 Hz. The RMSD is almost an order of magnitude smaller than the total chemical shift difference between experimental repeats. This indicates that the differential relaxation rates gave rise to very broad ^{15}N-anti-TROSY linewidths, such that Equation 4.19 cannot fully account for the error in peak position for anti-TROSY lines even with SNR > 20 observed for all residues in IP sub-spectrum and for most residues in AP sub-spectrum.

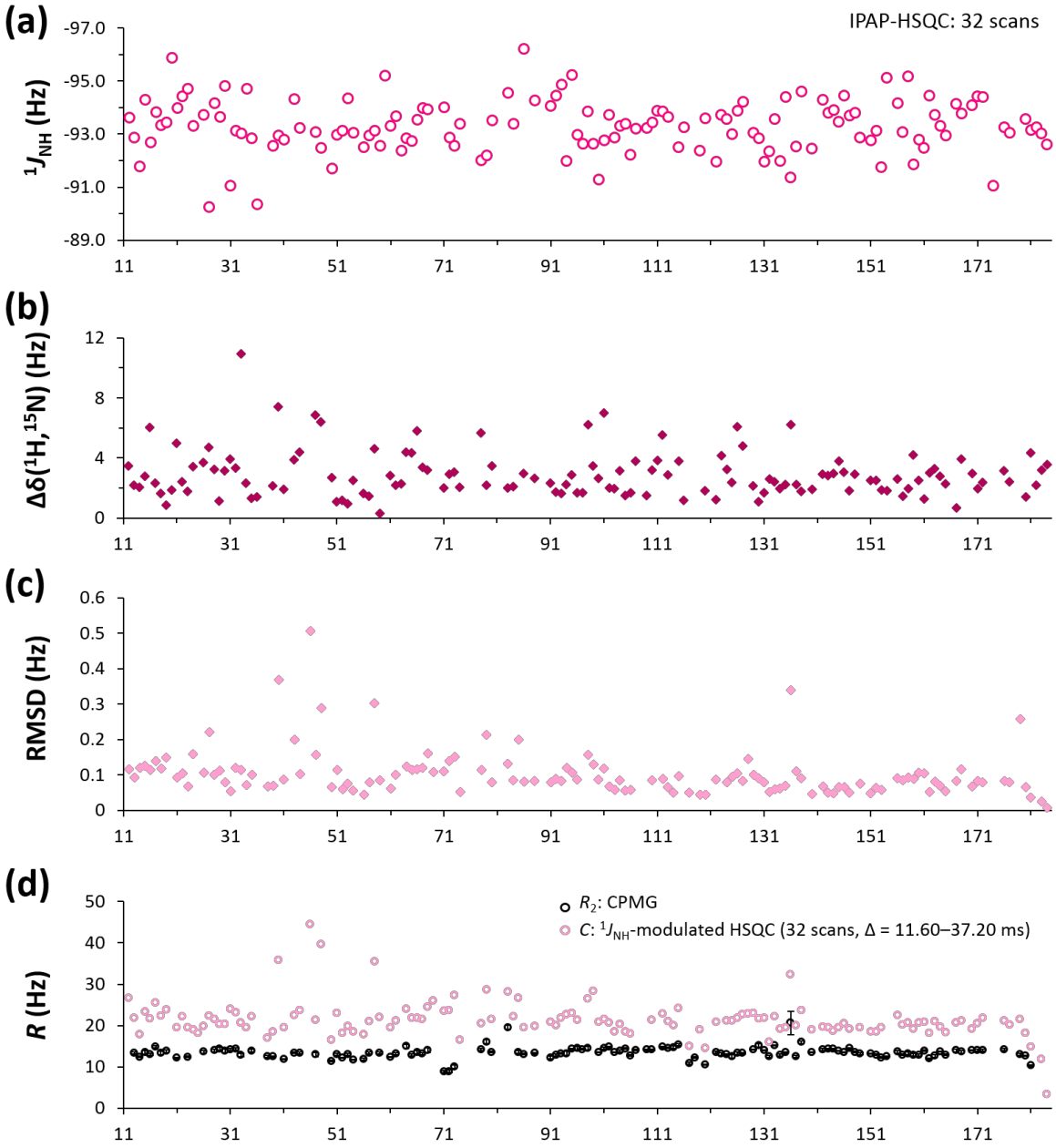


Figure 4.7: Example of $^1J_{NH}$ for RalA and the associated errors obtained from IPAP-HSQC
(a) $^1J_{NH}$ from IPAP-HSQC (32 scans). **(b)** Experimental error for $^1J_{NH}$ is calculated as the sum of chemical shift difference for each component of the ^{15}N doublet. **(c)** $^1J_{NH}$ RMSD value for a residue was calculated as the sum of RMSD for the position of each peak in the ^{15}N doublet. Peak position RMSD was calculated as $RMSD_{pos} = \{1/SNR\ (0.26/t_{1max} + 0.067\ R_2)\}$, where R_2 was taken as the decay rate (C) from the sinusoidal decay function fitted to $^1J_{NH}$-modulated HSQC (32 scans, Δ = 11.60–37.20 ms). Residues without error bars were due to missing C value. **(d)** Transverse relaxation rate constants measured by Dr. Arooj Shafiq using CPMG [374] or by fitting $^1J_{NH}$-modulated HSQC peak intensities with sinusoidal decay function.

To identify H-bonds, Δ^1J_{NH} values were calculated as the difference between $^1J_{NH}$ of RalAΔC-GDP and random coil values reported by Xiang *et al*. [201]. Δ^1J_{NH} values from IPAP-HSQC (32 scans) were plotted against the secondary structure and coloured by value on the crystal structure (Figure 4.8). Similar results were obtained for other implementations of IPAP-HSQC. Taking the reported error of ± 0.35 Hz [201], residues with $\Delta^1J_{NH} < –0.35$ Hz are therefore hypothesised to be H-bonded while residues with $\Delta^1J_{NH} > 0.35$ Hz are not H-bonded. Overall, the Δ^1J_{NH} values showed good correlation with the H-bonding patterns determined by WhatIf. Residues with Δ^1J_{NH} values < –0.35 Hz are mostly located in the central regions of α-helices and β-strands. Residues with $\Delta^1J_{NH} > 0.35$ Hz tend to be at the ends of α-helices and in loops.

Several residues showed mismatches against backbone-to-backbone H-bond predictions by WhatIf. Most of the mismatches could be accounted for once H-bonds from backbone amides to side-chains (as in the approach by Xiang *et al.*) and GDP were considered. The validity of this approach requires further assessment of the geometry and strength of the H-bond in question, as scalar coupling is mediated through electron cloud polarisation. Nevertheless, several Δ^1J_{NH} values still showed persistent mismatches against WhatIf predictions (Figure 4.8c). Such residues are mostly located at the ends of helices, ends of β-strands, loops and tight turns. While these mismatches are hypothesised to largely reflect the large experimental errors in IPAP-HSQC experiments, they might also be due to slight differences in the single conformation captured by the crystal structure and the ensemble of slightly different conformations in an aqueous protein sample. Since backbone amides with more internal motion of the N-H vector might have weaker H-bonds which lead to mismatch against WhatIf prediction, attempts were made to correlate these mismatches against the order parameter determined from ^{15}N relaxation experiments (R_1, R_2 and heteronuclear NOE) [374]. The order parameter predicts spatial restriction of the N-H vector and has values between 0 (unrestricted internal motion) and 1 (no internal motion). However, no correlation can be found between Δ^1J_{NH} mismatch and low order parameter values of < 0.85, thus indicating that the mismatches are most likely due to large experimental errors.

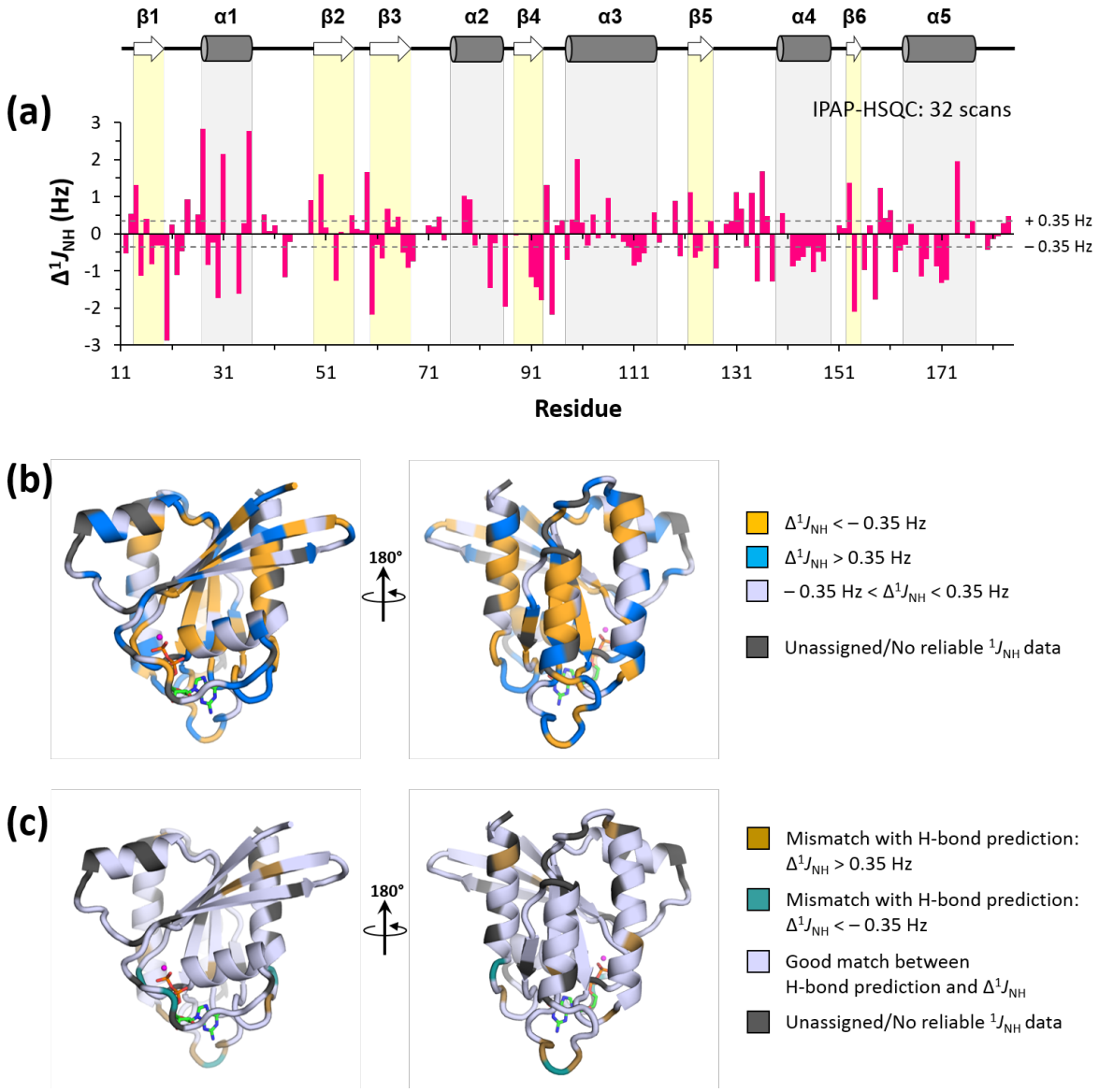


Figure 4.8: Example of Δ^1J_{NH} values for RalAΔC-GDP from IPAP-HSQC
(a) Δ^1J_{NH} determined by IPAP-HSQC (32 scans) are plotted against residue number and secondary structure. The grey dotted lines indicate the ± 0.35 Hz error range for the amino acid-specific random coil values reported by Xiang *et al* [201]. Several residues did not yield reliable data due to lack of assignments or were excluded from analysis due to peak overlap. **(b)** Δ^1J_{NH} values are mapped onto the crystal structure of RalAΔC-GDP and coloured by value. Δ^1J_{NH} < –0.35 Hz (yellow) is predictive of the presence of H-bond; Δ^1J_{NH} > 0.35 Hz predicts the lack of H-bond; –0.35 Hz < Δ^1J_{NH} < 0.35 Hz is within the reported error range and H-bond presence cannot be definitively determined. **(c)** Coloured residues show mismatch between presence/absence of H-bonds determined from Δ^1J_{NH} values vs. prediction by WhatIf.

Different implementations of Quantitative *J* modulation experiments were recorded to evaluate the impacts of SNR and lineshape on the measured $^1J_{NH}$ values (Table 4.3).

Quantitative *J* modulation implementation		**No. (%) of residues assigned**	**No. (%) of residues with ≥ 13 reliable Δ periods**
$^1J_{NH}$-modulated HSQC	8 scans, Δ = 11.60–37.20 ms	133 (79%)	124 (73%)
	16* scans, Δ = 11.60–37.20 ms	135 (80%)	132 (78%)
	32 scans, Δ = 11.60–37.20 ms	133 (79%)	128 (76%)
	32 scans, Δ = 32.88–58.48 ms	133 (79%)	116 (69%)
	8* scans, Δ = 32.88–58.48 ms	133 (79%)	104 (62%)
$^1J_{NH}$-modulated TROSY-HSQC	64* scans, Δ = 32.88–58.48 ms	129 (76%)	101 (60%)
	48* scans, Δ = 22.24–47.84 ms	130 (77%)	117 (69%)

Table 4.3: Residues with reliable $^1J_{NH}$ measurements in different $^1J_{NH}$-modulated HSQC implementations
* indicates that different number of scans were recorded for different Δ periods (referred to as "scan multiplier"). All experiments were recorded at 298 K on a 500 MHz spectrometer.

SNR was varied by recording experiments with different numbers of scans (Figure 4.9 insets). Since the signal intensity is modulated as an exponentially-decaying sinusoidal function of the *J* evolution period (Δ), the SNR at certain Δ are small and might lead to inaccuracies in the fitted peak intensities and errors in $^1J_{NH}$ measurements. Several implementations were attempted with different number of scans being recorded for different Δ periods (referred to as "scan multiplier"; see Table 4.4) in order to yield comparable SNR across all Δ. The effect of the scan multiplier is illustrated in Figure 4.9. The effects of lineshape on $^1J_{NH}$ measurements were evaluated by recording experiments in the absence vs. presence of TROSY effects. The lengths of Δ periods were also varied to examine how experimental error might be influenced by the averaging of relaxation rates of in-phase and antiphase operators at longer Δ.

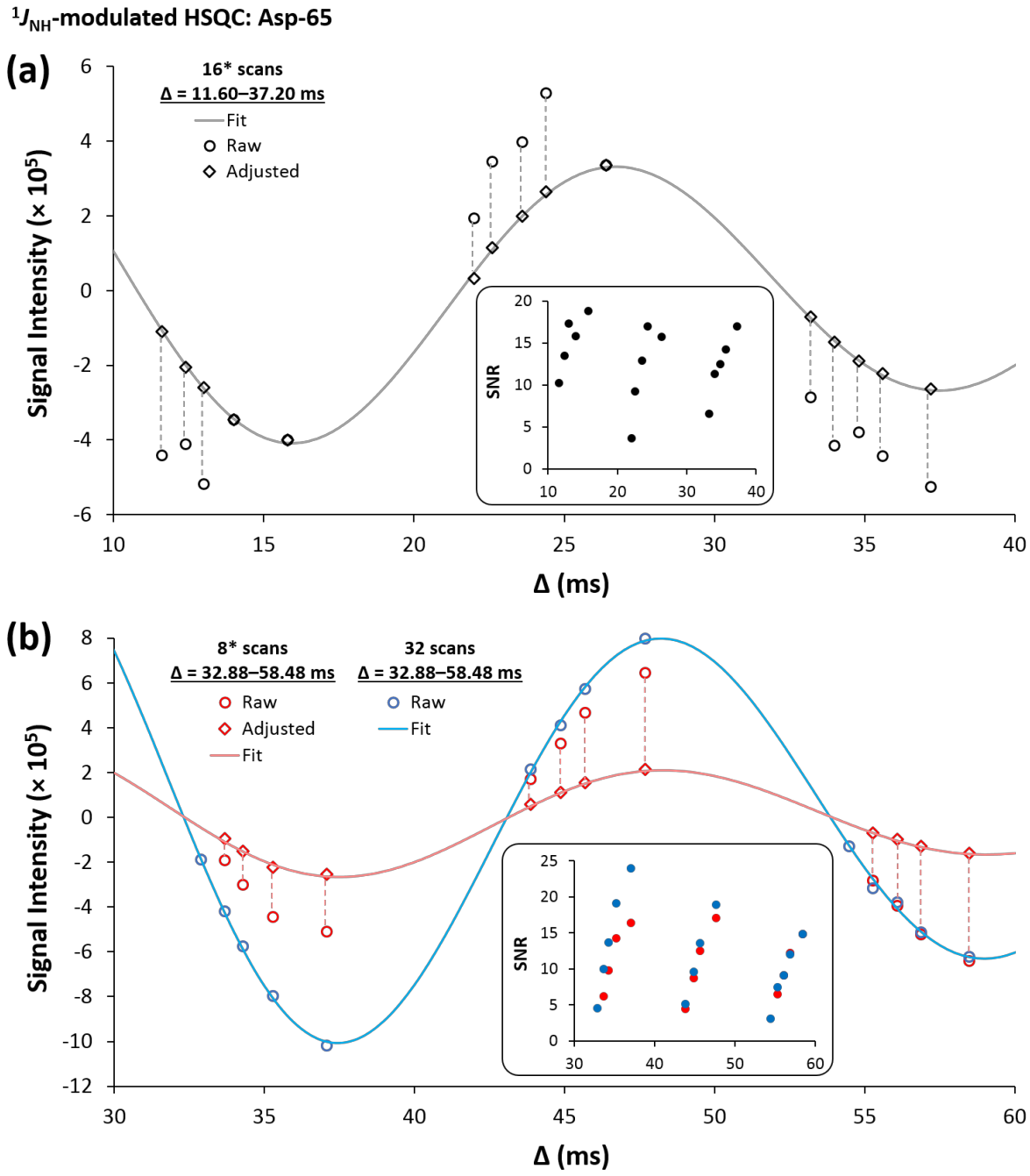


Figure 4.9: Example of the effect of the scan multiplier on peak intensities
The raw intensities (with scan multiplier), adjusted intensities (effect of scan multiplier taken away), fitted decaying sinusoidal function and raw SNR (insets) are shown for Asp-65 of RalAΔC-GDP in $^1J_{NH}$-modulated HSQC recorded with **(a)** 16* scans, Δ = 11.60–37.20 ms, and **(b)** 8* scans, Δ = 32.88–58.48 ms vs. 32 scans, Δ = 32.88–58.48 ms.

$^1J_{NH}$-modulated HSQC						$^1J_{NH}$-modulated TROSY-HSQC					
16* scans Δ = 11.60–37.20 ms			**8* scans Δ = 32.88–58.48 ms**			**64* scans Δ = 32.88–58.48 ms**			**48* scans Δ = 22.24–47.84 ms**		
Δ (ms)	Multiplier	No. Scans	Δ (ms)	Multiplier	No. Scans	Δ (ms)	Multiplier	No. Scans	Δ (ms)	Multiplier	No. Scans
11.60	4	64	32.88	2	16	32.88	2	128	22.24	2	96
12.40	2	32	33.68	2	16	33.68	2	128	23.04	2	96
13.00	2	32	34.28	2	16	34.28	2	128	23.64	2	96
14.00	1	16	35.28	2	16	35.28	2	128	24.64	2	96
15.80	1	16	37.08	2	16	37.08	2	128	26.44	2	96
22.00	6	96	43.28	3	24	43.28	3	192	32.64	2	96
22.60	3	48	43.88	3	24	43.88	3	192	33.24	2	96
23.60	2	32	44.88	3	24	44.88	3	192	34.24	2	96
24.40	2	32	45.68	3	24	45.68	3	192	35.04	2	96
26.40	1	16	47.68	3	24	47.68	3	192	37.04	2	96
33.20	4	64	54.48	4	32	54.48	4	256	43.84	3	144
34.00	3	48	55.28	4	32	55.28	4	256	44.64	3	144
34.80	2	32	56.08	4	32	56.08	4	256	45.44	3	144
35.60	2	32	56.88	4	32	56.88	4	256	46.24	3	144
37.20	2	32	58.48	4	32	58.48	4	256	47.84	3	144

Table 4.4: Scans recorded for each Δ period in $^1J_{NH}$ modulation experiments with scan multiplier

Stringent criteria must be employed to maximise the accuracy of fitting peak intensities as a function of Δ. The analysis therefore only includes residues with unambiguous and non-overlapping resonance assignments, and peak intensities determined for at least 13 values of Δ spanning two zero-crossings on the intensity modulation graph. These selection criteria permitted the fitting of $^1J_{NH}$ for 60–78% of the 169 assignable residues for different implementations of $^1J_{NH}$-modulated HSQC, hence providing similar coverage as IPAP-HSQC experiments (Table 4.3). Figure 4.10 shows that $^1J_{NH}$-modulated HSQC (16* scans, Δ = 11.60–37.20 ms) yielded the greatest fraction of residues with reliable peak intensities in ≥ 13 values of Δ. Increasing Δ to 22.24–47.84 ms or 32.88–58.48 ms gave decreased number of residues with reliable peak intensities across ≥ 13 values of Δ.

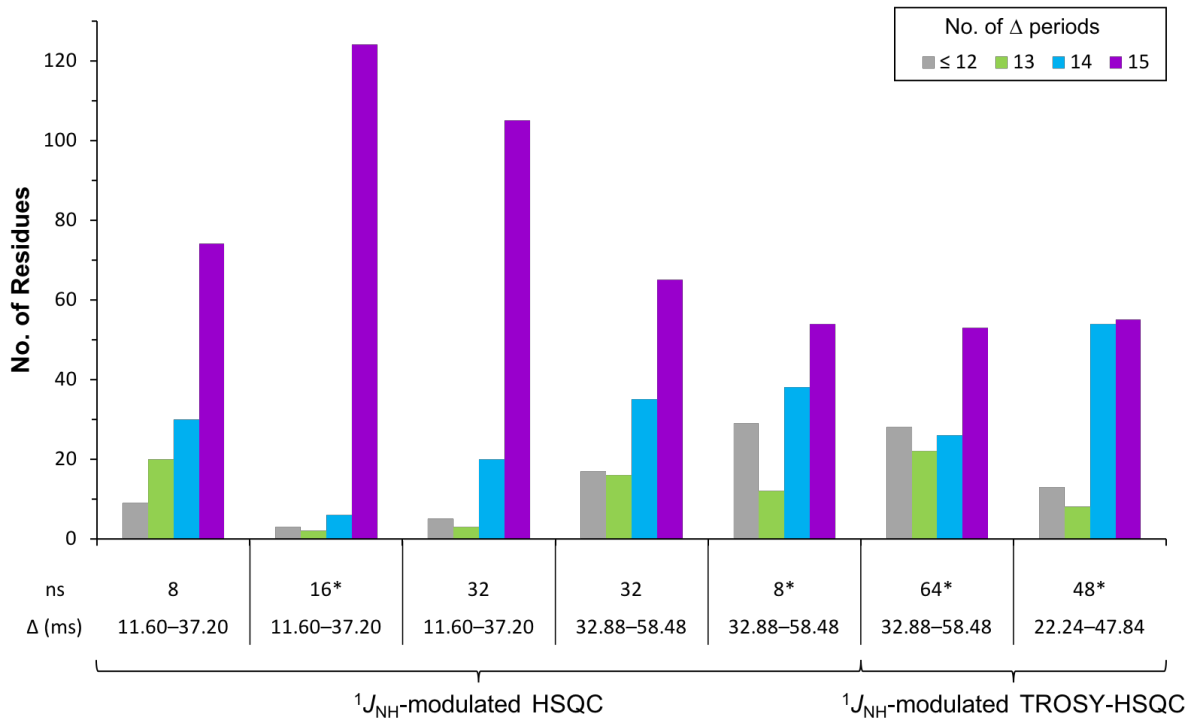


Figure 4.10: Histogram showing distribution of residues with reliable peak intensities for different numbers of Δ periods for different implementations of $^1J_{NH}$-modulated HSQC
All experiments were recorded at 298 K on a 500 MHz spectrometer.

In $^1J_{NH}$-modulated HSQC (8 scans, Δ = 11.60–37.20 ms), $^1J_{NH}$ ranged from –91 to –95 Hz while Δ^1J_{NH} span from –1.8 to + 2.5 Hz (Figure 4.11). These ranges appear to be narrower than those observed from different implementations of IPAP-HSQC.

The precision of IPAP-HSQC and $^1J_{NH}$-modulated HSQC measurements were evaluated, and examples are shown in Figure 4.12. Figure 4.12a shows cross comparison of Δ^1J_{NH} from two experimental repeats of IPAP-HSQC (32 scans), showing that several residues were outside the ± 0.35 Hz error range (grey dashed lines). Figure 4.12b shows comparison of Δ^1J_{NH} from $^1J_{NH}$-modulated HSQC (8* scans vs. 32 scans, Δ = 32.88–58.48 ms), where all residues were within the ± 0.35 Hz error range. Note that since direct experimental repeats of $^1J_{NH}$-modulated HSQC were not recorded, data from $^1J_{NH}$-modulated HSQC (8* scans, Δ = 32.88–58.48 ms) were compared against those from $^1J_{NH}$-modulated HSQC (32 scans, Δ = 32.88–58.48 ms). Since longer Δ periods yield fewer data points for intensity modulation

fitting (Figure 4.10), the longest Δ periods of 32.88–58.48 ms were chosen to estimate the upper limit of experimental error.

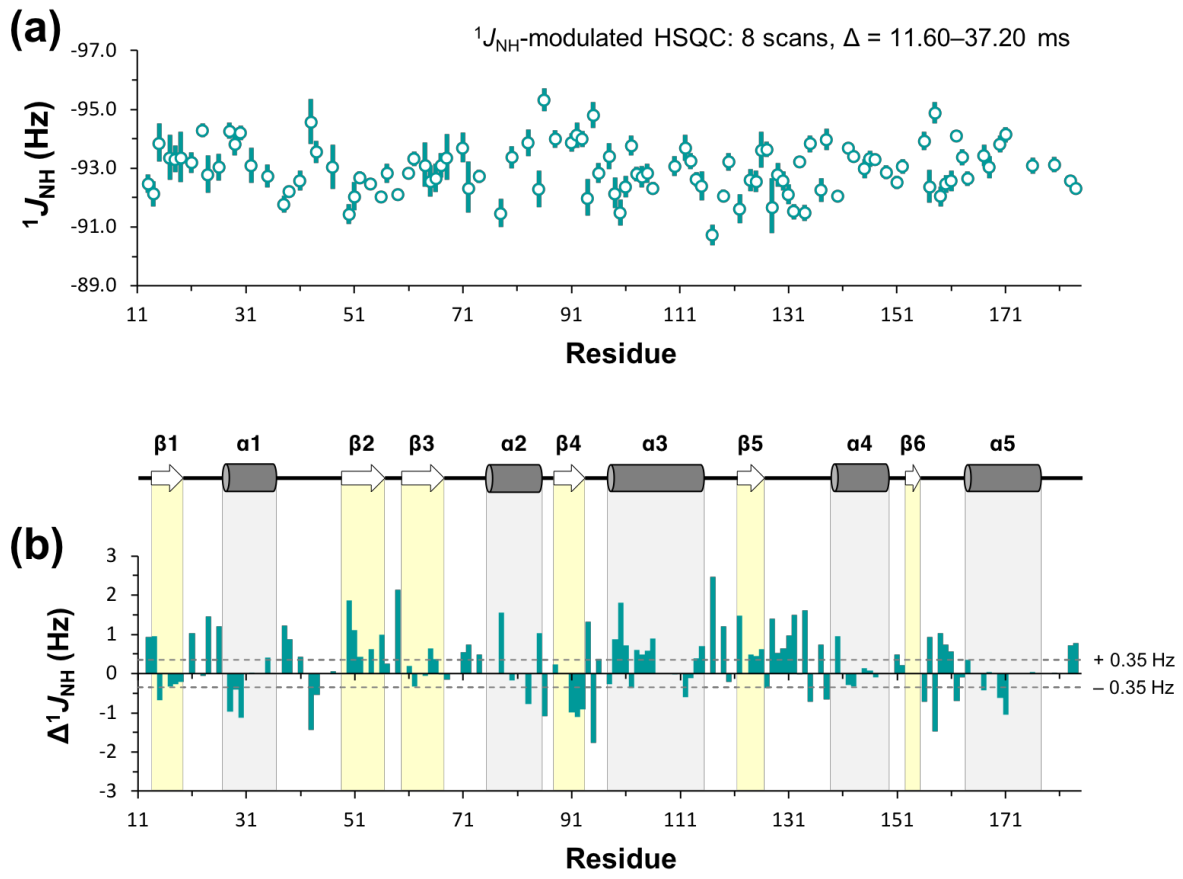


Figure 4.11: Example of $^1J_{NH}$ and Δ^1J_{NH} values for RalAΔC-GDP from $^1J_{NH}$-modulated HSQC
(a) $^1J_{NH}$ and **(b)** Δ^1J_{NH} determined by $^1J_{NH}$-modulated HSQC (8 scans, Δ = 11.60–37.20 ms) are plotted against residue number and secondary structure. In panel a, the vertical lines show the range between the 2.5% and 97.5% confidence limits of the $^1J_{NH}$ values obtained from fitting of peak intensities with the sinusoidal decay function. In panel b, the grey dotted lines indicate the ± 0.35 Hz error range for the amino acid-specific random coil values reported by Xiang *et al* [201]. Several residues did not yield reliable data due to lack of assignments or were excluded from analysis due to peak overlap or there were < 13 reliable peak intensities across different Δ periods.

To cross-compare the precision across all experimental repeats and different implementations of IPAP-HSQC and $^1J_{NH}$-modulated HSQC, the differences in residue-specific Δ^1J_{NH} between different experiments, referred to as $\Delta\Delta^1J_{NH}$, are summarised using box plots. Examples of $\Delta\Delta^1J_{NH}$ box plots are shown in Figure 4.12c, showing that the range of $\Delta\Delta^1J_{NH}$ values between experimental repeats of IPAP-HSQC and SOFAST-IPAP-HSQC are similar, but are consistently bigger than $\Delta\Delta^1J_{NH}$ values between different implementations of

$^1J_{NH}$-modulated HSQC. This indicates that to first-order approximation, $^1J_{NH}$-modulated HSQC shows higher precision than IPAP-HSQC.

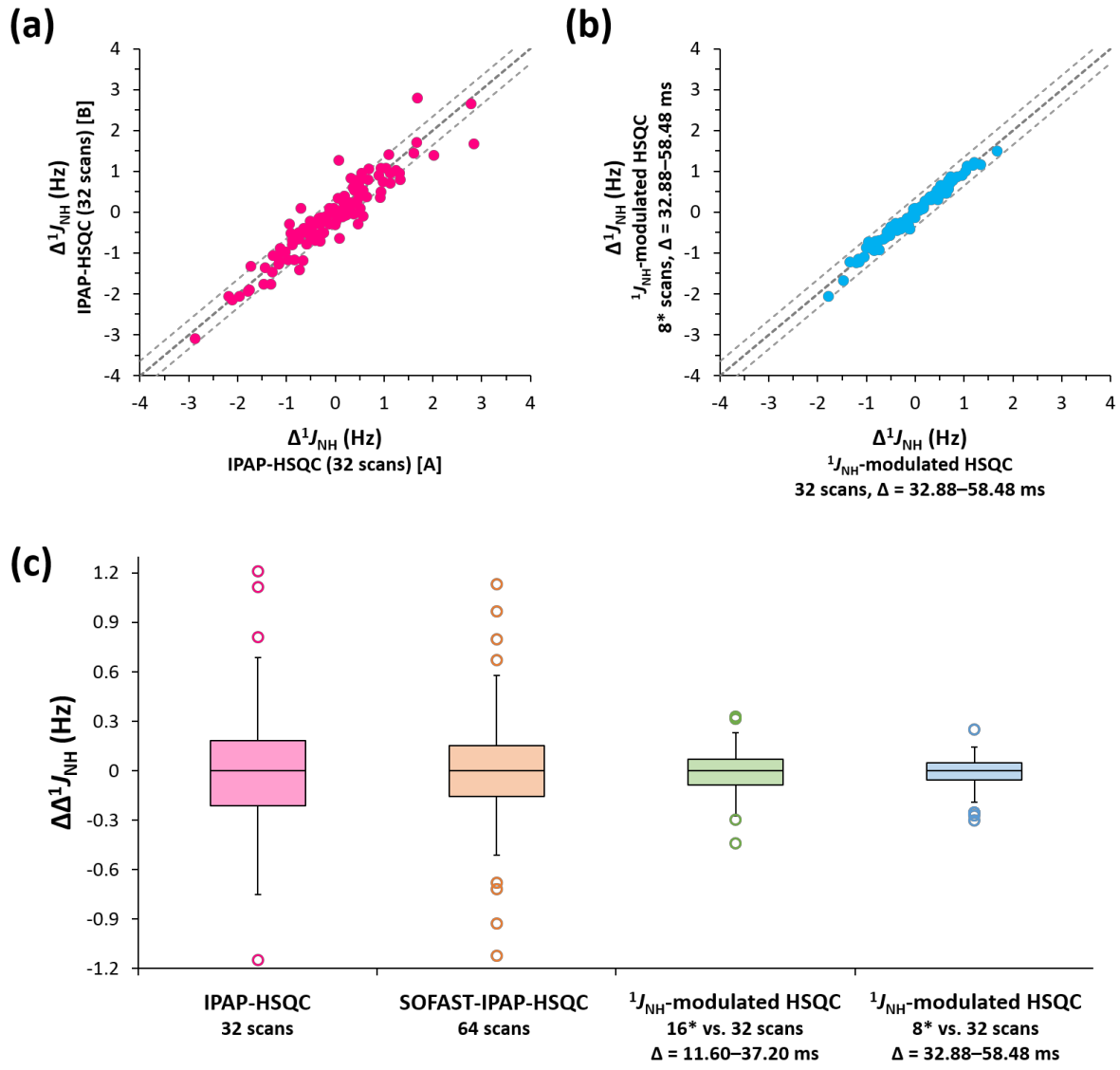


Figure 4.12: Comparison of the precision of Δ^1J_{NH} values across experimental repeats and different implementations of IPAP-HSQC and $^1J_{NH}$-modulated HSQC
(a) Comparison of Δ^1J_{NH} between two experimental repeats (denoted [A] and [B]) of IPAP-HSQC (32 scans). **(b)** Comparison of Δ^1J_{NH} from $^1J_{NH}$-modulated HSQC (8* scans vs. 32 scans, Δ = 32.88–58.48 ms). In panels a and b, the coordinates of each data point represent residue-specific Δ^1J_{NH} of the specified experimental repeats/implementations. The expected linear correlation with gradient = 1 is shown as bold grey dotted lines. The light grey dotted lines represent the ± 0.35 Hz error range reported by Xiang *et al* [201]. **(c)** Box plots showing the interquartile range and outliers in $\Delta\Delta^1J_{NH}$ (differences in residue-specific Δ^1J_{NH}) between different experimental repeats/implementations.

Δ^1J_{NH} values from $^1J_{NH}$-modulated HSQC appear to show consistent positive offset compared to those from IPAP-HSQC, as exemplified in Figure 4.13a. Pairwise comparison of the average $\Delta\Delta^1J_{NH}$ between different IPAP-HSQC and $^1J_{NH}$-modulated HSQC experiments is shown in Figure 4.13b. Pairwise comparison of different IPAP-HSQC experiments (top left corner) shows that on average, residue-specific Δ^1J_{NH} values are within 0.05 Hz between different experimental repeats and implementations. Pairwise comparison of different IPAP-HSQC and $^1J_{NH}$-modulated HSQC experiments (top right) shows that on average, Δ^1J_{NH} values are 0.22–0.39 Hz greater than those observed in IPAP-HSQC. For example, the Δ^1J_{NH} values from $^1J_{NH}$-modulated HSQC (8 scans, Δ = 11.60–37.20 ms) are on average 0.36 Hz greater than those from IPAP-HSQC (32 scans), providing a semi-quantitative measure of the positive offset shown in Figure 4.13a. Given the good correlation between Δ^1J_{NH} values from IPAP-HSQC experiments and H-bonds predicted by WhatIf, the positive offset in Δ^1J_{NH} from $^1J_{NH}$-modulated HSQC experiments indicates the presence of systematic error. Lengthening the Δ period appears to decrease the systematic error, although it remains rather big. This is because the transverse magnetisation interconverts more times between in-phase and antiphase operators, and the effective decay rates for the in-phase and antiphase operators become closer to the average relaxation rate. The change in systematic error with the length of the Δ period likely explains the varied average $\Delta\Delta^1J_{NH}$ values in pairwise comparison of different $^1J_{NH}$-modulated HSQC implementations (bottom right corner).

$^1J_{NH}$-modulated TROSY-HSQC (48* scans, Δ = 22.24–47.84 ms) yielded the smallest systematic errors. This suggests that the low SNR at long Δ also contributed towards experimental error, and a better balance between SNR and averaging of differential relaxation rates can be achieved with Δ = 22.24–47.84 ms. Other parameters including number of scans, the scan multiplier and TROSY implementation appeared to have minimal effect ($\Delta\Delta^1J_{NH} \leq 0.05$ Hz) on the systematic error.

Δ^1J_{NH} from $^1J_{NH}$-modulated TROSY-HSQC (48* scans, Δ = 22.24–47.84 ms) are mapped onto the crystal structure (Figure 4.14a), showing some correlation with H-bonds predicted by WhatIf. Most residues with mismatches against H-bond prediction showed $\Delta^1J_{NH} > 0.35$ Hz but were predicted to be non-H-bonded, despite H-bonds from backbone

amides to side-chains and GDP having been taken into account (Figure 4.14b). This illustrates the presence of a positive systematic error in Δ^1J_{NH} values.

The systematic error was crudely corrected by subtracting 0.23 Hz from all experimentally-determined Δ^1J_{NH} values. 0.23 Hz reflects the average difference in Δ^1J_{NH} between $^1J_{NH}$-modulated TROSY-HSQC (48* scans, Δ = 22.24–47.84 ms) and the IPAP-HSQC experiments. This approach yielded better correlation with H-bonds predicted by WhatIf (Figure 4.14c–d).

Summarising the proof-of-concept studies on RalAΔC-GDP, $^1J_{NH}$ measurements by IPAP-HSQC give good correlations with H-bonds predicted by WhatIf but are plagued by large experimental errors attributable to broad linewidths in the AP spectrum due to differential relaxation. On the other hand, $^1J_{NH}$-modulated HSQC give small experimental deviations but yield big systematic errors which can only be partially reduced by increasing the duration of the *J* evolution period, provided that the concomitant decrease in SNR does not lead to increase in experimental error.

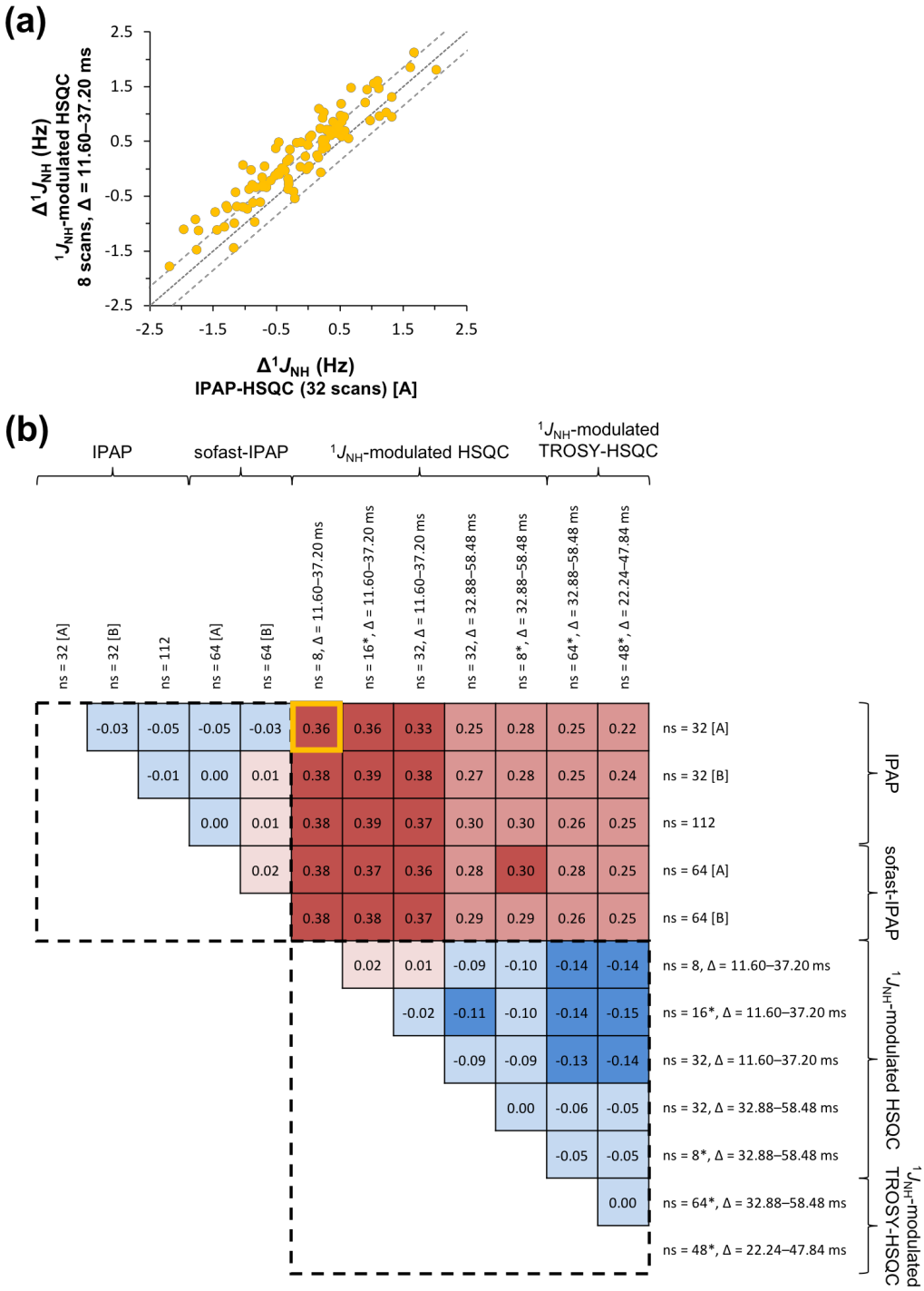


Figure 4.13: Average pairwise difference in Δ^1J_{NH} across different IPAP-HSQC and $^1J_{NH}$-modulated HSQC experiments

(a) Comparison of Δ^1J_{NH} from $^1J_{NH}$-modulated HSQC (8 scans, Δ = 11.60–37.20 ms) vs. IPAP-HSQC (32 scans). The grey lines represent the ± 0.35 Hz error range reported by Xiang *et al* [201]. **(b)** The average differences in Δ^1J_{NH} across all assigned residues for pairwise comparisons of different IPAP-HSQC and $^1J_{NH}$-modulated HSQC experiments. The yellow box shows the average pairwise difference in Δ^1J_{NH} illustrated in panel a.

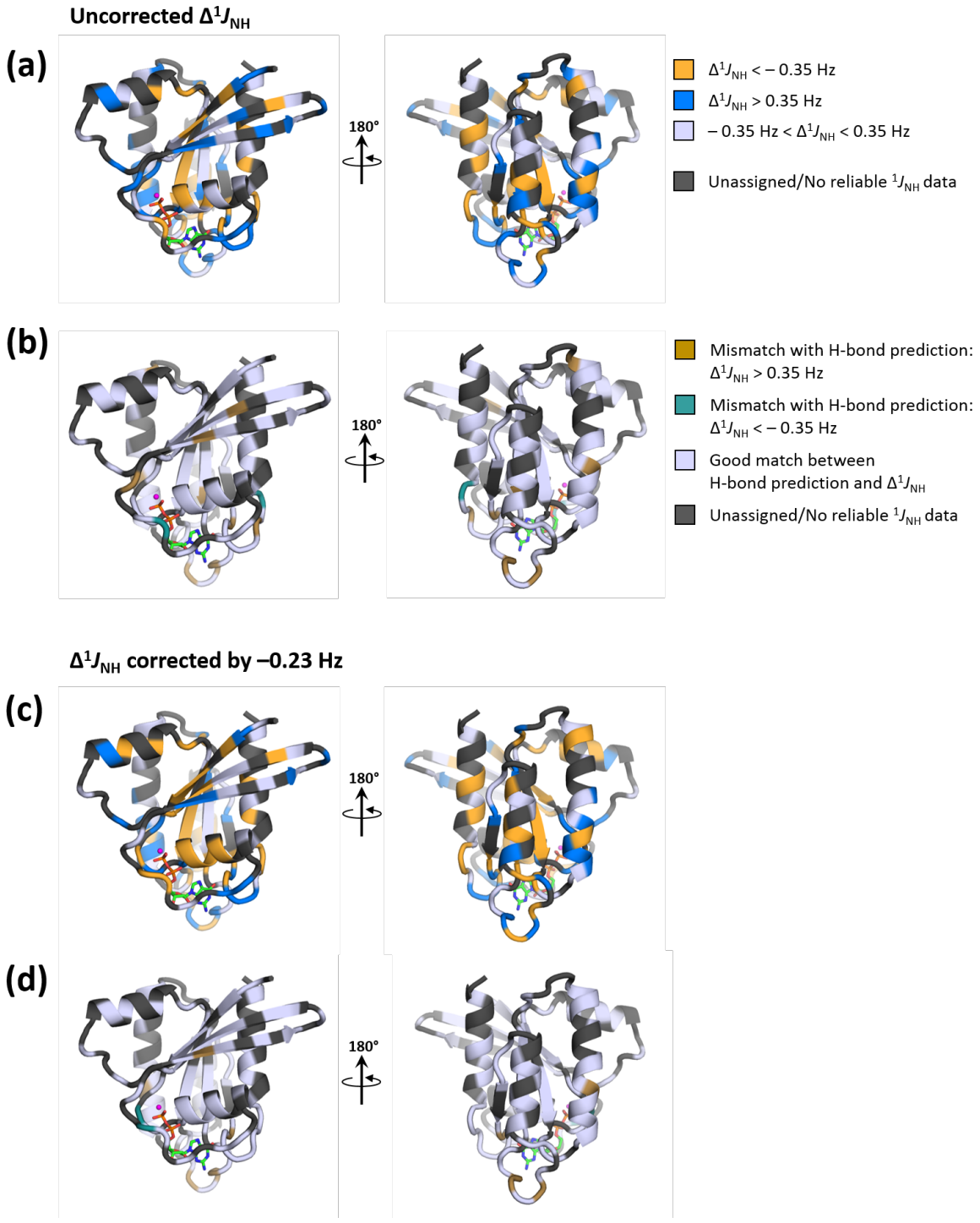


Figure 4.14: Example of Δ^1J_{NH} values for RalAΔC-GDP from $^1J_{NH}$-TROSY-HSQC
(a) Δ^1J_{NH} determined by $^1J_{NH}$-modulated TROSY-HSQC (48* scans, Δ = 22.24–47.84 ms) are mapped onto the crystal structure of RalAΔC-GDP and coloured by value. **(b)** Mismatches between Δ^1J_{NH} values and H-bond prediction by WhatIf. **(c)** Δ^1J_{NH} corrected by –0.23 Hz. **(d)** Mismatches against H-bond prediction by WhatIf after –0.23 Hz correction.

4.3.2 Expression, purification and NMR of OmpX

^{15}N-labelled OmpX was expressed in *E. coli* as inclusion bodies and purified using published methods [377] (Figures 4.15 & 4.16a–b), before being refolded by dropwise rapid dilution of the unfolded protein in the presence of detergent (DPC) and 0.5 M L-arginine-hydrochloride (see Materials and Methods). The denaturant was removed by several overnight dialyses [285]. The refolding efficiency was assessed by SDS-PAGE, where the folded species and the misfolded species migrated as separate bands on the gel (Figure 4.16c). The final sample contained ~ 90% ^{15}N-labelled OmpX (995 µM, in 20 mM sodium phosphate pH 6.8, 100 mM NaCl, 5 mM EDTA, 3.8% DPC, 90% H_2O/10% D_2O).

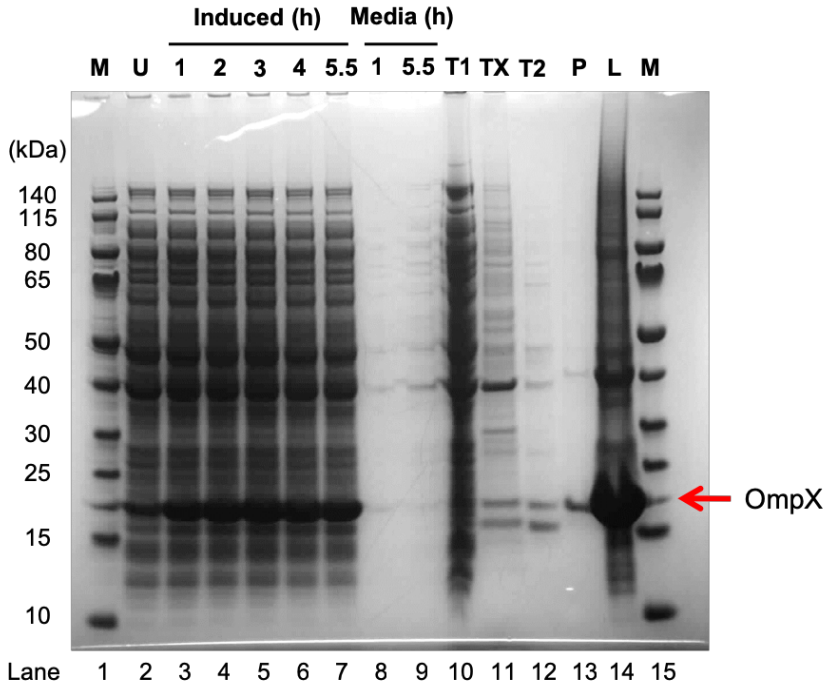


Figure 4.15: Expression and purification of ^{15}N-labelled OmpX
Representative SDS-PAGE gel showing the expression and purification of ^{15}N-OmpX. M: molecular weight marker; U: whole cell lysate of un-induced BL21(DE3); Induced 1 – 5.5: whole cell lysate of BL21(DE3) at 1, 2, 3, 4, 5.5 h post-induction; Media 1 + 5.5: culture media at 1 and 5.5 h post-induction; T1: flow-through from first Buffer TE wash; TX: flow-through from Buffer TE + 2% (v/v) Triton X-100 wash; T2: flow-through from second Buffer TE wash; P: insoluble pellet in 6 M Gdm-Cl; L: sample loaded onto a size exclusion column.

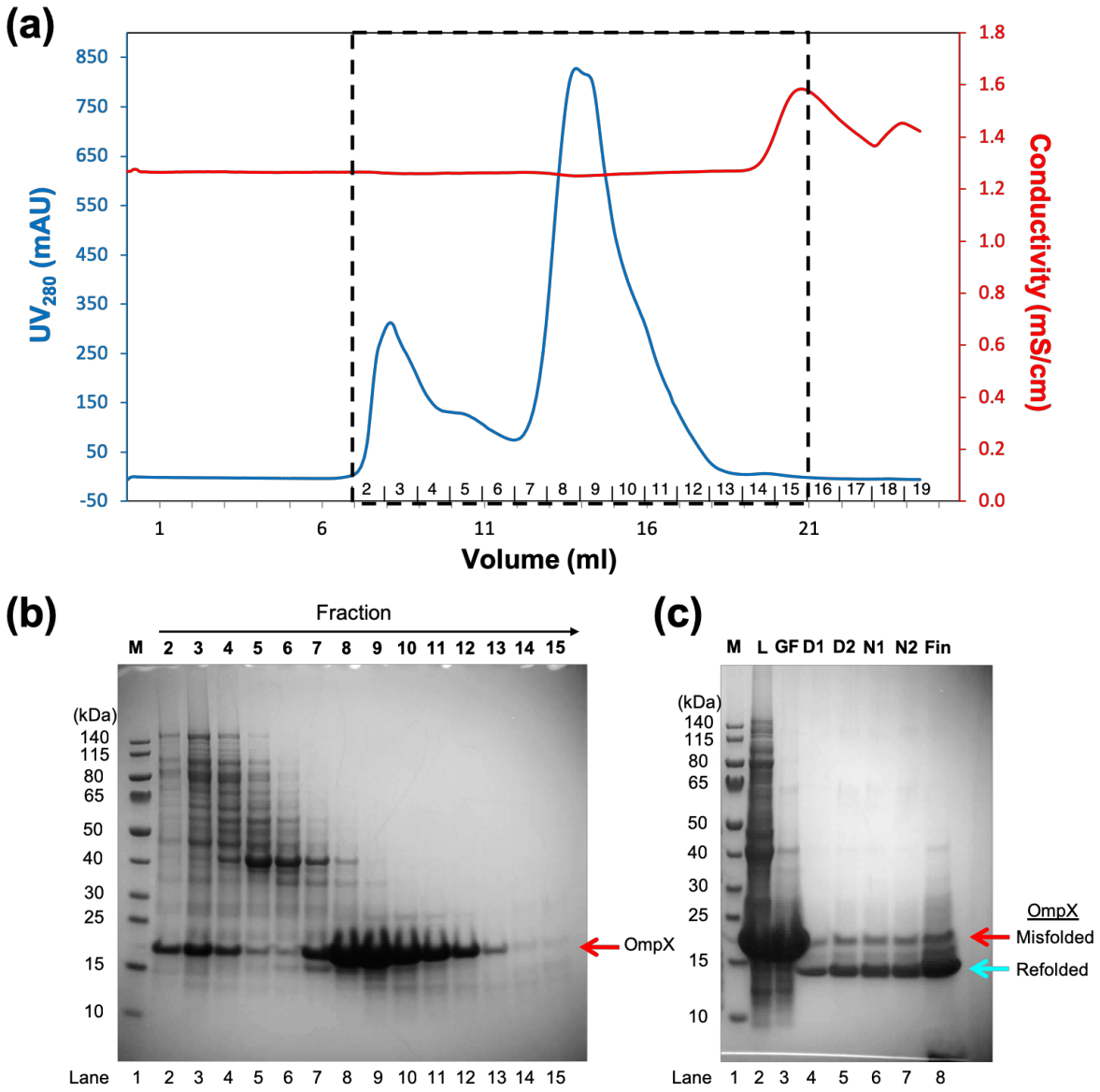


Figure 4.16: Size exclusion chromatography and refolding of ^{15}N-labelled OmpX
(a) Representative size exclusion chromatography trace for OmpX purification using a Superdex 200 10/300 GL column running at 0.12 ml/min. The volume of each eluted fraction is 1 ml. The total sample loaded onto the column is equivalent to 1 L of *E. coli* BL21(DE3) culture. **(b)** Representative SDS-PAGE gel for fractions 2−15 (boxed in panel a). OmpX (~ 16.5 kDa) is expected to migrate between 15 and 25 kDa (red arrow). **(c)** Representative SDS-PAGE gel for refolding of OmpX, overnight dialyses and concentration of the protein sample. Red and cyan arrows show different migrations for OmpX in its denatured (misfolded) conformation and refolded conformation, respectively. M: molecular weight marker; L: pre-gel filtration sample; GF: post-gel filtration sample; D1 & D2: samples after each round of overnight dialysis into Buffer D; N1 & N2: samples after each round of overnight dialysis into Buffer N; Fin: final ^{15}N-OmpX NMR sample (995 µM).

The quality of the sample was verified by comparing an acquired 2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC (Figure 4.17a) against a synthetic spectrum generated using the chemical shifts from BMRB entry 18797 [285]. Extra peaks in the acquired spectrum arise from side-chain NH_2 groups (not shown on the synthetic spectrum) or residual misfolded species within the sample. All backbone amide assignments can be unambiguously transferred from the synthetic spectrum to the acquired spectrum, with little differences in chemical shifts observed (Figure 4.17b). The sample remained stable at room temperature for more than 2 months, with no observable changes in the HSQC spectrum over this time period.

A 3D ^{15}N-separated NOESY-HSQC spectrum was recorded, and the NOE cross peak positions were used to confirm the BMRB deposited (BMRB 18797) resonance assignments. The presence of a large number of sequential, inter-β-strand and long-range H^N-H^N NOE interactions (Figure 4.18) was observed, confirming the integrity of the tertiary protein structure. Several intra-strand/intra-loop NOEs were also observed, especially in the loop regions of OmpX.

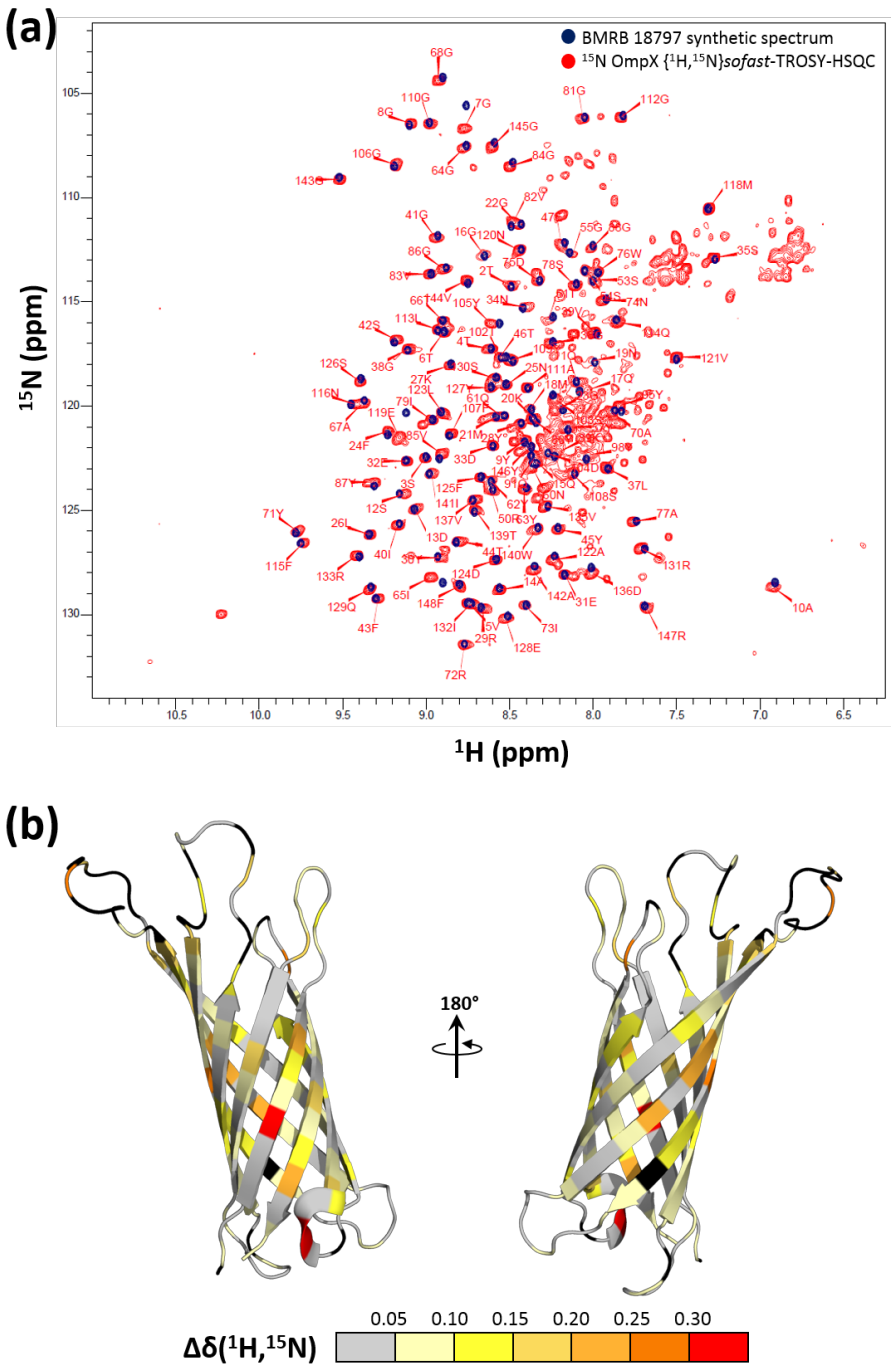


Figure 4.17: 2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC spectrum of ^{15}N-OmpX
(a) Overlay of a 2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC 800 MHz spectrum recorded at 318 K against a synthetic spectrum generated using the chemical shifts from BMRB entry 18797 [285]. **(b)** Chemical shift differences $\Delta\delta(^{1}H,^{15}N) = [(\Delta\delta^{1}H)^{2} + (\Delta\delta^{15}N)^{2}/6]^{0.5}$ in panel a are mapped onto the solution-state NMR structure of OmpX (PDB 2M07).

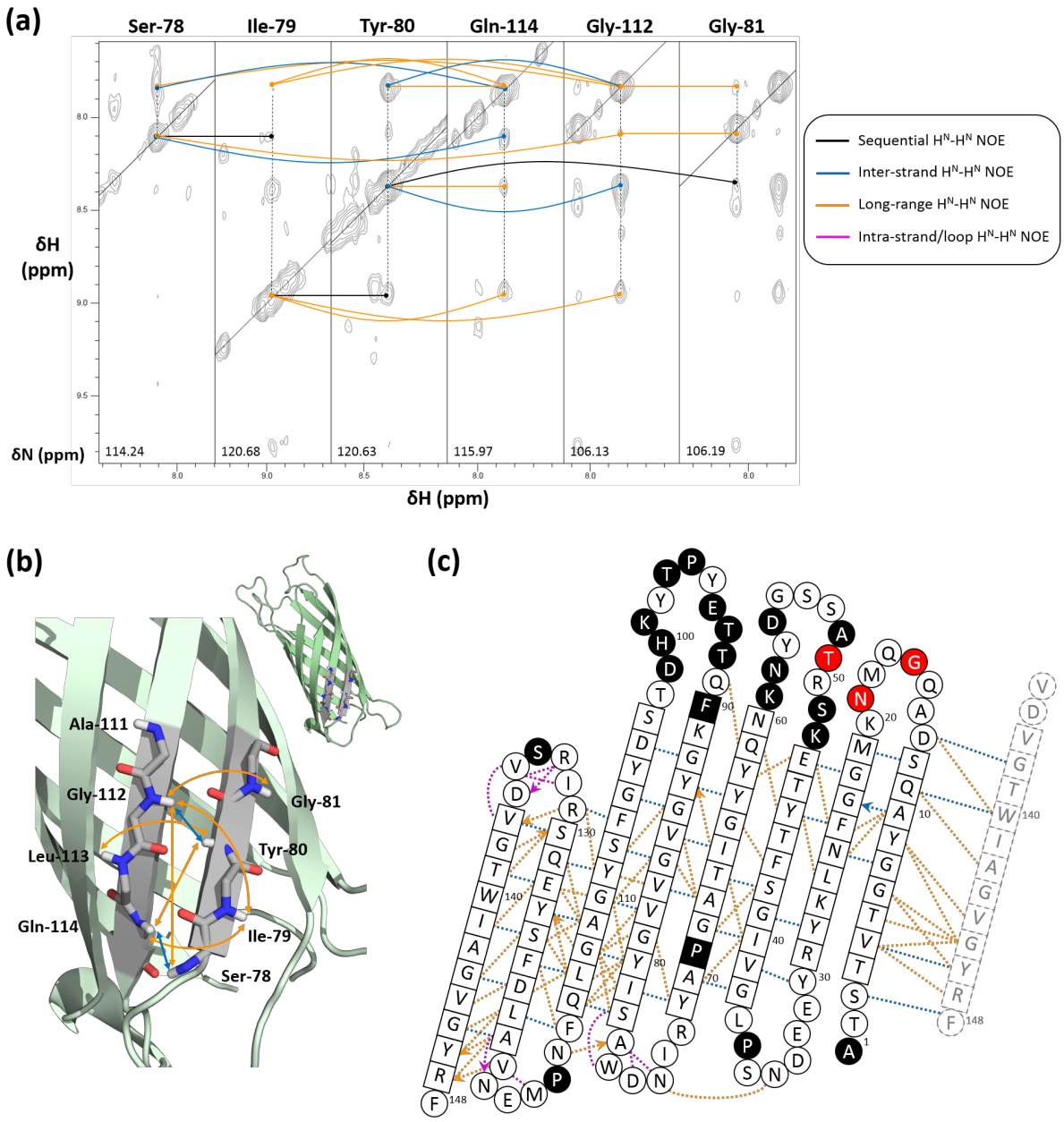


Figure 4.18: Backbone amide NOE interactions in OmpX
Backbone amide NOEs are shown as coloured lines. Sequential NOEs are shown in black, inter-strand NOEs in blue, long-range NOEs in orange, and intra-strand/intra-loop NOEs in magenta. **(a)** ^{1}H-^{1}H cross sections from a 3D ^{15}N-separated NOESY-HSQC spectrum of ^{15}N-OmpX taken at the ^{15}N frequencies of residues 78−81, 112 and 114, showing sequential, inter-strand and long-range NOEs between these residues. **(b)** Inter-strand and long-range NOEs shown in panel a are mapped onto the solution-state NMR structure of OmpX (PDB 2M07). **(c)** Backbone amide NOE interactions are mapped onto a topology plot of OmpX. Coloured lines with arrowheads indicate the direction of the one-way NOE observed, whilst coloured lines without arrowheads indicate the observation of both forward and reverse NOEs. Unassigned residues in BMRB 18797 are highlighted in black, and residues with no peaks observed in the ^{1}H-^{15}N plane of the acquired 3D ^{15}N-separated NOESY-HSQC spectrum are highlighted in red. Sequential HN-HN NOEs are omitted for clarity.

4.3.3 $^1J_{NH}$ of OmpX in DPC micelles

The proof-of-concept studies on RalAΔC have demonstrated the feasibility of measuring $^1J_{NH}$ for a 20.9 kDa soluble protein. Further studies were conducted using OmpX in DPC micelles to assess the feasibility of $^1J_{NH}$ measurements in folded membrane proteins. Four experiments were recorded:

1. $^1J_{NH}$-modulated HSQC (40 scans, Δ = 11.60–37.20 ms)
2. $^1J_{NH}$-modulated HSQC (96 scans, Δ = 11.60–37.20 ms)
3. $^1J_{NH}$-modulated TROSY-HSQC (40 scans, Δ = 11.60–37.20 ms)
4. SOFAST-IPAP-HSQC (64 scans)

The number and fraction of residues with $^1J_{NH}$ measurements out of 128 assignable residues are evaluated in Table 4.5. All $^1J_{NH}$-modulated HSQC experiments were recorded with Δ = 11.60–37.20 ms, as the large OmpX-DPC protein-detergent complex of 37.2 kDa will have fast transverse relaxation rates, yielding very low SNR at long Δ periods.

Experiment		No. (%) of residues assigned	No. (%) of residues with ≥ 10 reliable Δ periods	No. (%) of residues with ≥ 13 reliable Δ periods
$^1J_{NH}$-modulated HSQC	40 scans	125 (98%)	88 (69%)	44 (34%)
	96 scans	124 (97%)	90 (70%)	51 (40%)
$^1J_{NH}$-modulated TROSY-HSQC	40 scans	125 (98%)	95 (74%)	58 (45%)
SOFAST-IPAP-HSQC	64 scans	112 (88%)	N/A	N/A

Table 4.5: Residues with reliable $^1J_{NH}$ measurements in $^1J_{NH}$-modulated HSQC and IPAP-HSQC implementations
All experiments were recorded at 318 K on a 800 MHz spectrometer.

Similar to the studies on RalAΔC, overlapping resonances were excluded from analysis to minimise error. Out of the 128 assignable backbone amide resonances [285], $^1J_{NH}$ values were obtained for 88% residues in the SOFAST-IPAP-HSQC experiment, and ≥ 97% residues in different implementations of $^1J_{NH}$-modulated HSQC. However, subsequent selection of residues with measured peak intensities from ≥ 13 Δ periods in $^1J_{NH}$-modulated HSQC would yield low protein coverage of only 34–45% (Table 4.5). Hence, a more lenient

set of criteria was used to enable comparable coverage (~70%) as those obtained in the studies on RalAΔC: residues which were analysed have unambiguous assignments and peak intensities from ≥ 10 Δ periods spanning two zero-crossings on the intensity modulation graph, with at least three data points on either side of each zero-crossing.

The measured $^{1}J_{NH}$ values ranged from −90.6 to −95.5 Hz for $^{1}J_{NH}$ modulation-based experiments, and from −86.7 to −98.2 Hz for the SOFAST-IPAP-HSQC experiment (Figure 4.19).

$\Delta^{1}J_{NH}$ ranged from −2.5 Hz to +2.5 Hz for $^{1}J_{NH}$ modulation-based experiments, whereas a much larger range of about −5.0 Hz to +6.0 Hz was observed in the SOFAST-IPAP-HSQC experiment (Figure 4.20). The larger range of $\Delta^{1}J_{NH}$ indicates large experimental errors due to broad linewidths in the AP spectrum, similar to previous observations with RalAΔC.

Figure 4.21a shows there is poor correlation between the expected H-bonding pattern based on $\Delta^{1}J_{NH}$ from different experiments and the H-bonding pattern predicted using WhatIf. Quite a few residues which are located in the middle of the β-barrel and are expected to be H-bonded show $\Delta^{1}J_{NH}$ > 0.35 Hz (Figures 4.21b–c).

Pairwise comparison of $\Delta^{1}J_{NH}$ values from different $^{1}J_{NH}$-modulated HSQC and SOFAST-IPAP-HSQC implementations against those from $^{1}J_{NH}$-modulated HSQC (40 scans) show poor correlation (Figure 4.22). The observed errors are also much larger than the ± 0.35 Hz error range reported by Xiang *et al* [201].

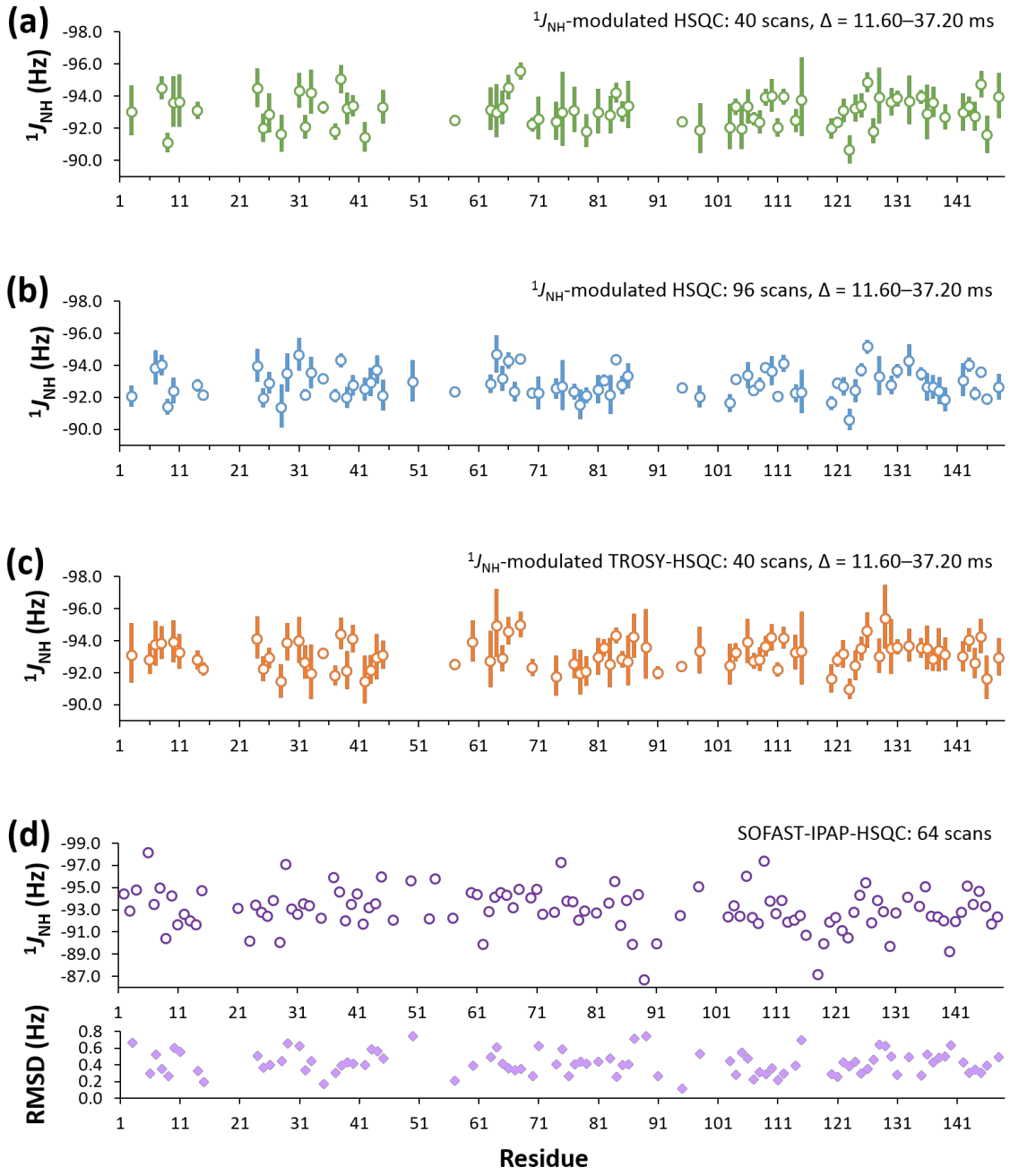


Figure 4.19: $^1J_{NH}$ for OmpX-DPC
(a–c) $^1J_{NH}$ from $^1J_{NH}$-modulated HSQC (40 scans) **(a)**, $^1J_{NH}$-modulated HSQC (96 scans) **(b)** and $^1J_{NH}$-modulated TROSY-HSQC (40 scans) **(c)**. All implementations were recorded with Δ = 11.60–37.20 ms. The vertical lines show the range between the 2.5% and 97.5% confidence limits of the $^1J_{NH}$ values obtained from fitting of peak intensities with the sinusoidal decay function. **(d)** Top panel: $^1J_{NH}$ from SOFAST-IPAP-HSQC (64 scans). Bottom panel: $^1J_{NH}$ RMSD, calculated as the sum of RMSD for peak position for each component of the ^{15}N doublet. Residues without RMSD are due to missing information on the transverse relaxation rate derived from the three $^1J_{NH}$-modulated HSQC experiments (no peak detected or inadequate fitting of intensity modulation). All experiments were recorded at 318 K on a 800 MHz spectrometer.

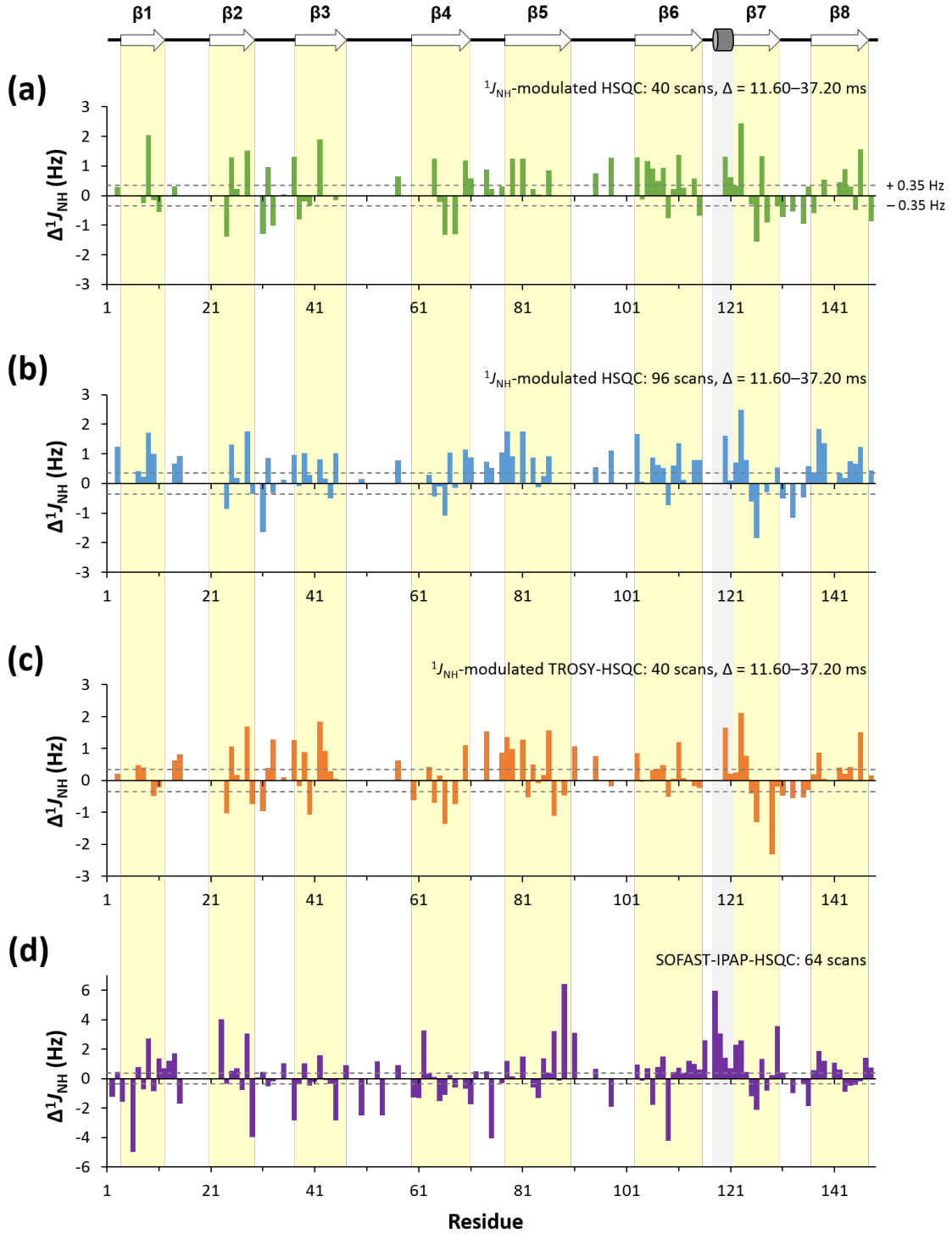


Figure 4.20: Δ^1J_{NH} for OmpX-DPC
Δ^1J_{NH} are plotted against residue number and secondary structure. The horizontal grey dotted lines indicate the ± 0.35 Hz error range for the amino acid-specific random coil values reported by Xiang *et al* [201]. Several residues did not yield reliable data due to lack of assignments or were excluded from analysis due to peak overlap or there were < 10 reliable peak intensities across different Δ periods.

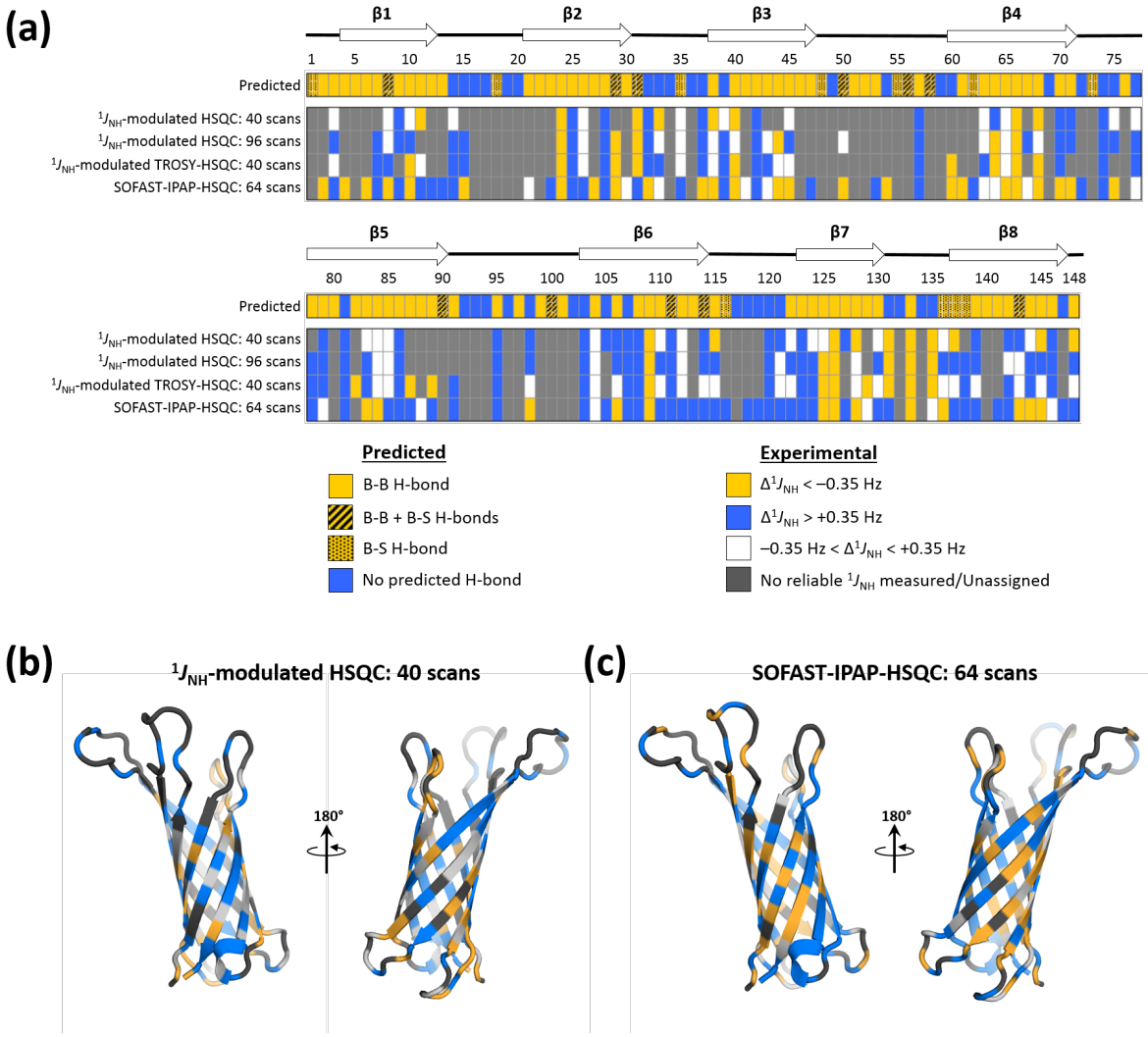


Figure 4.21: Comparison of experimental Δ^1J_{NH} values for OmpX-DPC against H-bond predictions using the WhatIf server

(a) Each residue of OmpX is colour-coded according to the H-bond predictions using WhatIf (Predicted), and Δ^1J_{NH} from different implementations of $^1J_{NH}$-modulated HSQC and SOFAST-IPAP-HSQC. B–B: backbone-to-backbone H-bond; B–S: backbone-to-sidechain H-bond. Δ^1J_{NH} from $^1J_{NH}$-modulated HSQC (40 scans) **(b)** and SOFAST-IPAP-HSQC (64 scans) **(c)** are mapped onto the solution-state NMR structure of OmpX (PDB 2M07) and coloured by value.

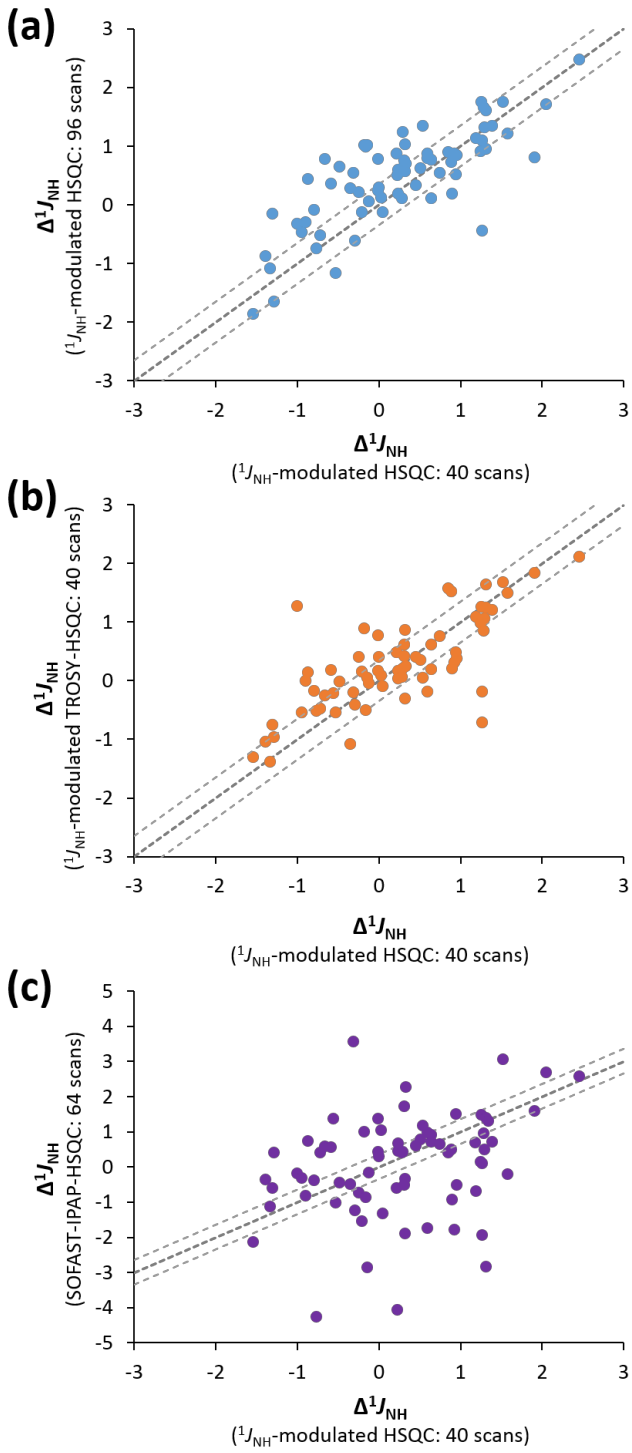


Figure 4.22: Poor correlation between Δ^1J_{NH} values derived from different $^1J_{NH}$ experiments
Δ^1J_{NH} values from $^1J_{NH}$-modulated HSQC (40 scans) are compared against those from $^1J_{NH}$-modulated HSQC (96 scans) **(a)**, $^1J_{NH}$-modulated TROSY-HSQC (40 scans) **(b)**, and SOFAST-IPAP-HSQC (64 scans) **(c)**. The expected linear correlation with gradient = 1 is shown as bold grey dotted lines. The light grey dotted lines represent the ± 0.35 Hz error range reported by Xiang *et al* [201].

According to the classical theory of Ramsey [378], an isotropic spin-spin coupling constant (SSCC) is given as the sum of the diamagnetic spin orbit (DSO), paramagnetic spin orbit (PSO), Fermi contact (FC) and spin dipole (SD) terms:

$$\widehat{H}_{\mathrm{SSCC}} = \widehat{H}_{\mathrm{DSO}} + \widehat{H}_{\mathrm{PSO}} + \widehat{H}_{\mathrm{FC}} + \widehat{H}_{\mathrm{SD}} \tag{4.20}$$

One-bond scalar coupling is dominated by the FC term, which largely arises from an overlap between the *s*-orbital spin density of one nucleus and the coupled nucleus. The SD term arises from nuclear magnetic moments interacting via the magnetic field induced by the spins of surrounding electrons. The DSO and PSO terms arise from nuclear magnetic moments interacting via the magnetic field induced by the orbital motion of surrounding electrons [185]. In particular, the PSO term becomes significant with the presence of π-orbital electrons (e.g. from aromatic rings) in close proximity [378,379].

Xiang *et al*. have observed several residues in ubiquitin with Δ^1J_{NH} values which do not correlate well with H-bonding patterns predicted by WhatIf. These residues are mostly within 6 Å away from an aromatic ring, which perturbs $^1J_{\mathrm{NH}}$ measurements due to a significant PSO contribution. The distance cut-off employed to account for such deviations in $^1J_{\mathrm{NH}}$ appears to be arbitrary, with Xiang *et al*. having employed a 6 Å cut-off from aromatic side-chains for ubiquitin [201] whereas Ma *et al*. have employed a 4 Å cut-off for GlnBP [379].

The distance between each backbone amide of OmpX from the closest aromatic ring was determined using Pymol (Figure 4.23). Given the abundance of aromatic side-chains, application of a 6 Å cut-off would yield only 9 out of 148 residues for further analysis. However, there is evident inconsistency between Δ^1J_{NH} and H-bond prediction for these 9 remaining residues, suggesting that perturbations of $^1J_{\mathrm{NH}}$ measurements by proximal aromatic rings are small compared to experimental and systematic errors. In fact, a computational study on the relative contributions of the Ramsey terms to *J*-coupling measurement has shown that the H-bonded residue pair 4–65 in ubiquitin, which are within the 6 Å cut-off distance applied by Xiang *et al*., only has 1.31% contribution from the PSO term whilst the Fermi contact term dominates (98.11% contribution) the $^1J_{\mathrm{NH}}$ coupling [380].

In summary, the discrepancy in Δ^1J_{NH} values between the recorded experiments and WhatIf H-bond predictions likely reflect problems in the precision and accuracy of the measurements contributed by using a protonated sample of a large detergent-protein complex. In comparison, other factors which might also lead to perturbation of $^1J_{NH}$ measurements, e.g. close proximity of aromatic rings, play a smaller part in the existing data.

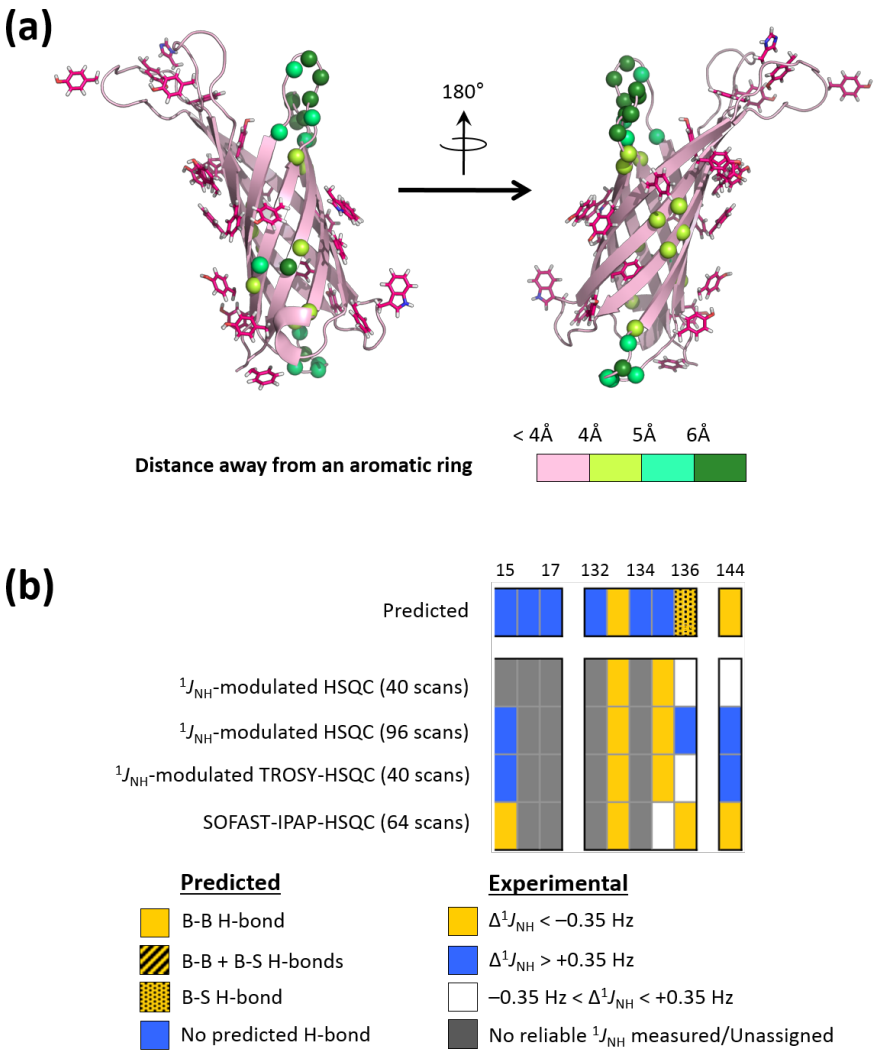


Figure 4.23: Effects of proximal aromatic rings on the Δ^1J_{NH} values in OmpX-DPC
(a) OmpX residues are colour-coded based on the distance of each backbone amide from the closest aromatic ring. Amide groups located beyond 4 Å, 5 Å and 6 Å away from an aromatic ring are represented using lime green, green and dark green spheres, respectively. Aromatic side-chains are shown using stick representation. **(b)** Residues remaining after application of a 6 Å distance cut-off are colour-coded according to the H-bond predictions using WhatIf (Predicted), and Δ^1J_{NH} from different implementations of $^1J_{NH}$-modulated HSQC and SOFAST-IPAP-HSQC. B–B: backbone-to-backbone H-bond; B–S: backbone-to-sidechain H-bond.

4.3.4 $^1J_{NH}$ of OmpX in urea

The NMR chemical shifts of unfolded OmpX in 8 M urea [91] and the structures of folded OmpX in different membrane mimetics [285] have been published, hence permitting facile assessment of the feasibility and experimental error of $^1J_{NH}$ measurements in both unfolded and folded membrane proteins.

To obtain OmpX in 8 M urea, OmpX in DPC was subjected to buffer exchange, gel filtration (Figure 4.24), dialysis and dilution/concentration with a 3 kDa MWCO centrifugal concentrator (see Materials and Methods). The final DPC concentration was determined to be < 0.02% by 1D ^{1}H NMR (Figure 4.25a). The resulting sample contained 221 μM ^{15}N-OmpX, in 8 M urea, 20 mM sodium phosphate pH 6.5, 0.1 mM NaN_3, 95% H_2O/5% D_2O, and yielded comparable 2D [^{1}H,^{15}N] TROSY-HSQC spectrum to the 2D [^{1}H,^{15}N] HSQC spectrum obtained by Tafer *et al* [91].

Given the aim of studying H-bond formation during protein folding by measuring $^1J_{NH}$, ^{15}N-OmpX in different urea concentrations was studied by NMR in order to access denatured states which sample folded conformations with increasing frequencies. However, precipitation of OmpX was observable by eye in < 6 M urea in the absence of detergent, thus limiting the range of this study to 6–8 M urea. Figure 4.26 shows the 2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC spectra recorded for ^{15}N-OmpX in 6 M, 7 M and 8 M urea. Addition of DPC was not considered, because of the large errors in Δ^1J_{NH} values for folded ^{15}N-OmpX attributable to the utilisation of a protonated sample containing a large detergent-protein complex (see Section 4.3.3).

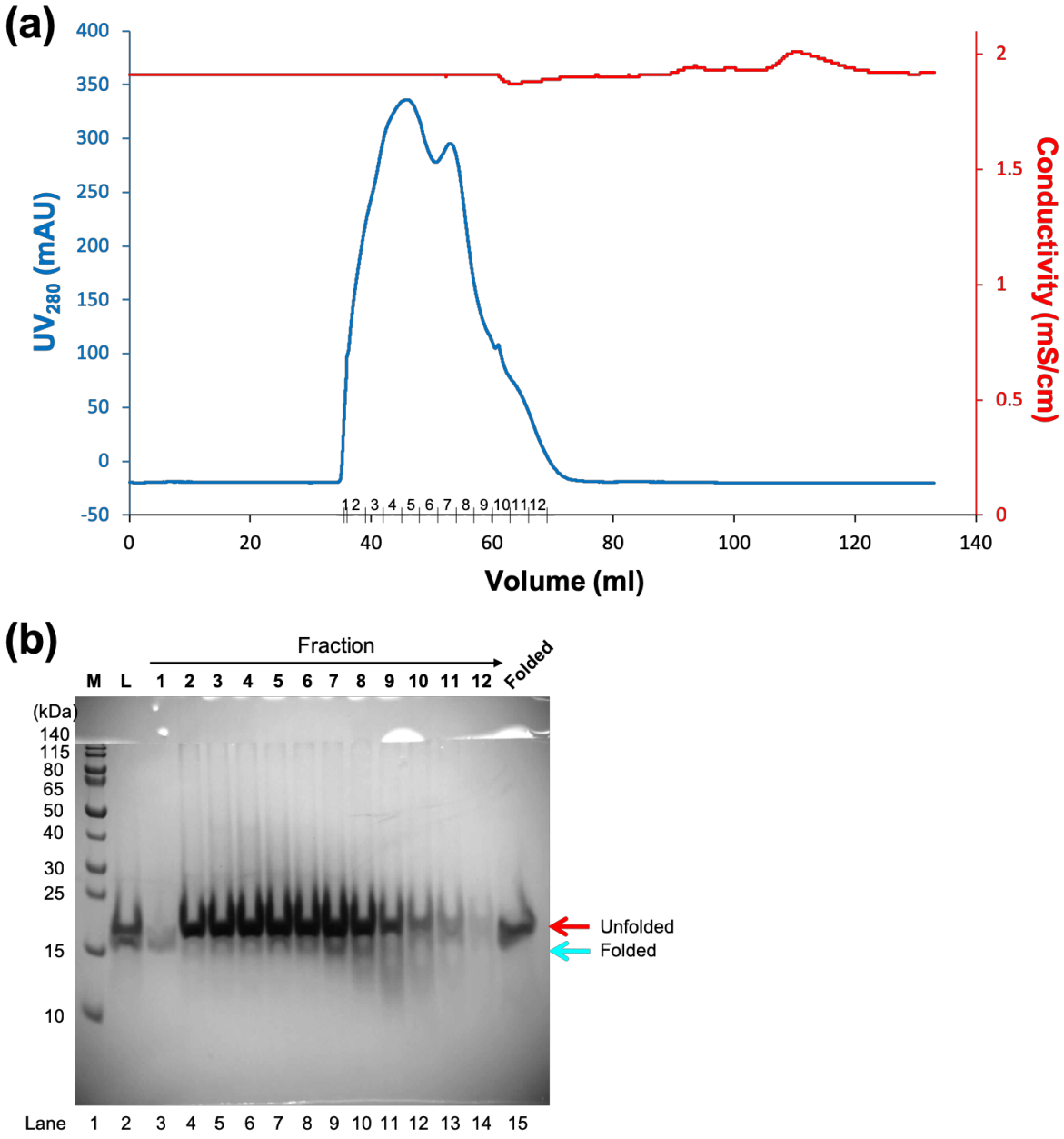


Figure 4.24: Size exclusion chromatography for exchange of ^{15}N-OmpX from DPC into 8 M urea

(a) Size exclusion chromatography trace for exchange of ^{15}N-OmpX into 8 M urea using a HiPrep 16/60 Sephacryl S-200 HR column running at 0.5 ml/min. The volume of each eluted fraction is 3 ml. **(b)** SDS-PAGE gel for fraction 1–12 from size exclusion chromatography. Red and cyan arrows show different migrations for OmpX in its unfolded conformation and folded conformation, respectively. M: molecular weight marker; L: pre-gel filtration sample; Folded: folded ^{15}N OmpX in DPC.

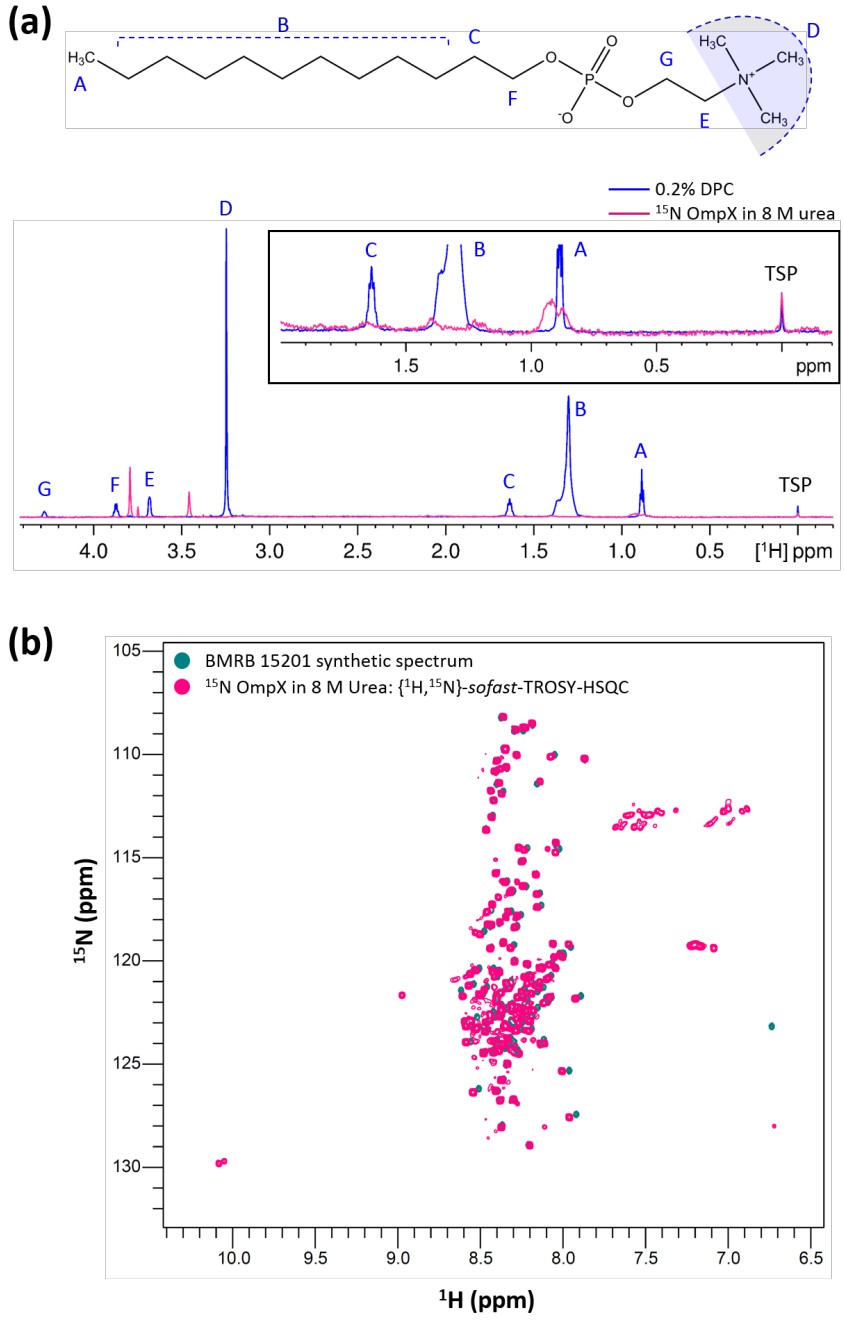


Figure 4.25: NMR spectra of ^{15}N-OmpX in 8 M urea
(a) 1D ^{1}H NMR (0–4 ppm region) of 0.2% DPC in 95% H_2O/5% D_2O (blue) (recorded at 293 K) and 221 µM ^{15}N-OmpX in 8 M urea (recorded at 288 K), showing no/negligible presence of DPC in the protein sample. Both samples contained TSP as internal reference. Both spectra were recorded with 4 scans and an acquisition time of 1 s, and processed with exponential window function and line-broadening of 2 Hz. Inset shows expanded view of the 0 – 2 ppm region, with the intensities of both spectra scaled according to the TSP peak. Chemical structure of DPC was drawn using ChemDraw (PerkinElmer). Assignments of DPC resonances are derived from [381]. **(b)** Overlay of a 2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC 800 MHz spectrum recorded at 288 K with 80 scans against a synthetic spectrum generated using the chemical shifts from BMRB entry 15201 [91].

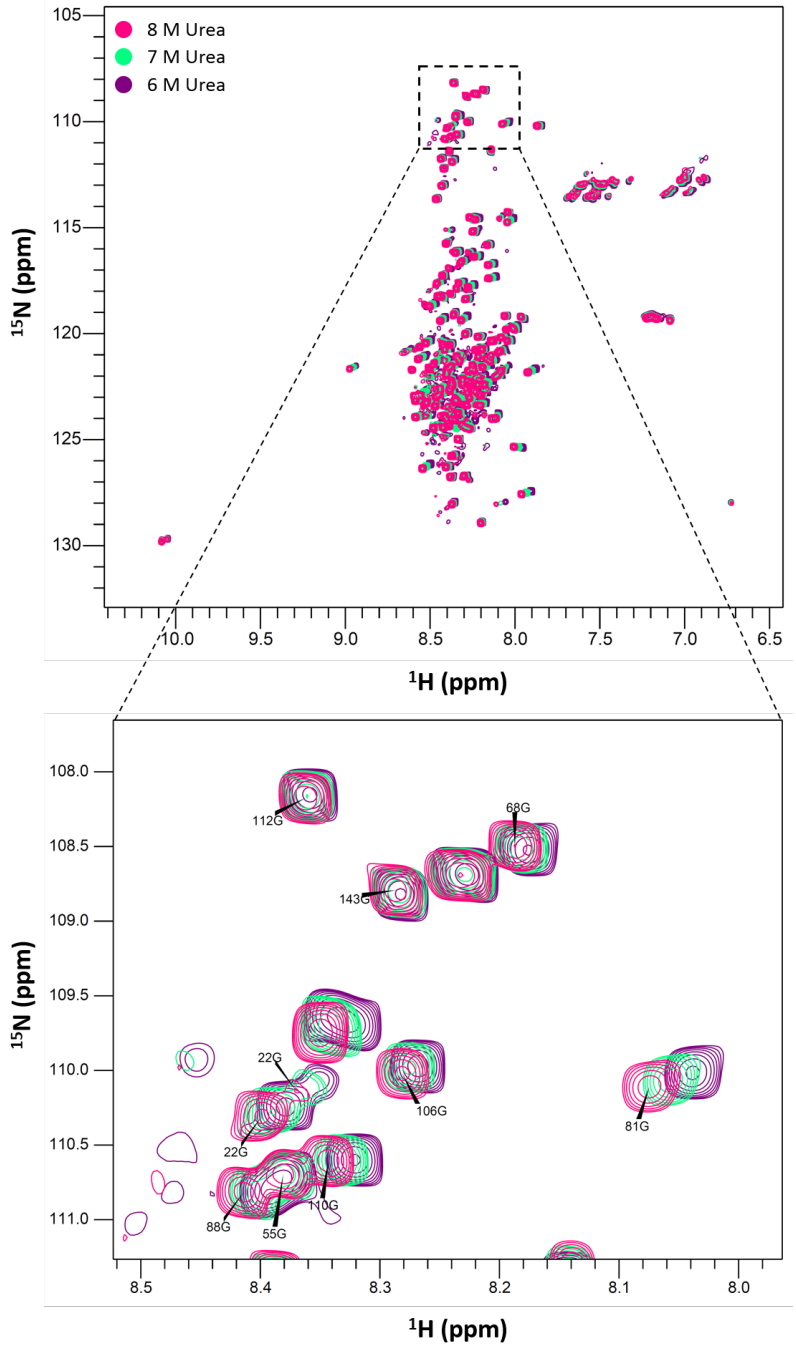


Figure 4.26: 2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC spectra of ^{15}N-OmpX in different urea concentrations.
166 µM ^{15}N-OmpX in 6 M urea, 97 µM ^{15}N-OmpX in 7 M urea, and 221 µM ^{15}N-OmpX in 8 M urea. All samples contained TSP as internal reference. All spectra were recorded on a 800 MHz spectrometer at 288 K and with 80 scans.

Three experiments were recorded, and $^{1}J_{NH}$ measurements were obtained for high fraction of residues out of the 144 reported assignable residues [91] (Table 4.6).

[Urea]	Experiment		No. (%) of residues assigned	No. (%) of residues with ≥ 13 reliable Δ periods
8 M	$^1J_{NH}$-modulated HSQC Δ = 22.24–47.84 ms	40 scans	131 (91%)	128 (89%)
	SOFAST-IPAP-HSQC	64 scans	126 (88%)	N/A
6 M	SOFAST-IPAP-HSQC	64 scans	124 (86%)	N/A

Table 4.6: Residues of OmpX in 8 M and 6 M urea with reliable $^1J_{NH}$ measurements
All experiments were recorded at 288 K on a 800 MHz spectrometer.

The measured $^1J_{NH}$ values ranged from −92.2 to −94.1 Hz across all three experiments (Figure 4.27). All Δ^1J_{NH} were > 0.35 Hz or were within the reported ± 0.35 Hz error range, with the exception of Gln-129, which has Δ^1J_{NH} = −0.44 Hz measured by SOFAST-IPAP-HSQC in the presence 6 M urea (Figure 4.28–4.29). Δ^1J_{NH} values of OmpX in 8 M urea from $^1J_{NH}$-modulated HSQC (40 scans) show good correlation with those from SOFAST-IPAP-HSQC (64 scans) (Figure 4.28d), indicating the absence of systematic errors which were observed for RalAΔC. Δ^1J_{NH} values of OmpX in 6 M urea showed a general negative offset against those of OmpX in 8 M urea. It is unknown whether this difference reflects a genuine global shift towards H-bonded conformations or an artifact due to different buffer conditions.

Based on Δ^1J_{NH} values, the lack of stable H-bonds for OmpX in 8 M urea is consistent with the amide proton exchange protection factors and structure calculations reported by Tafer *et al* [91]. 3D ^{15}N-resolved [^{1}H,^{1}H]-NOESY spectrum of OmpX in 8 M urea shows long-range NOEs in segments comprising residues 73–82 (I) and 137–145 (II). Structure calculations performed on segments I and II showed non-random local conformations. At least one of the two backbone H-bonds, Trp76 NH–O Ile73 and Ala77 NH–O Asn74, is present in 17 of the 20 conformers, of which 9 conformers show the presence of both H-bonds [91]. In 8 M urea, Δ^1J_{NH} of H-bond acceptors Ile-73 and Asn-74 were 0.41–0.51 Hz, and Δ^1J_{NH} of H-bond donor Ala-77 was slightly smaller (~ 0.20 Hz). Δ^1J_{NH} of Trp-76 could not be determined since the random coil $^1J_{NH}$ was not published by Xiang *et al* [201]. These results indicate that the backbone H-bonds determined by NMR structure calculations were below the detection limit of the methods for $^1J_{NH}$ measurement, due to the transient nature of the H-bonds and experimental errors which are expected to be larger than those reported by Xiang *et al* [201].

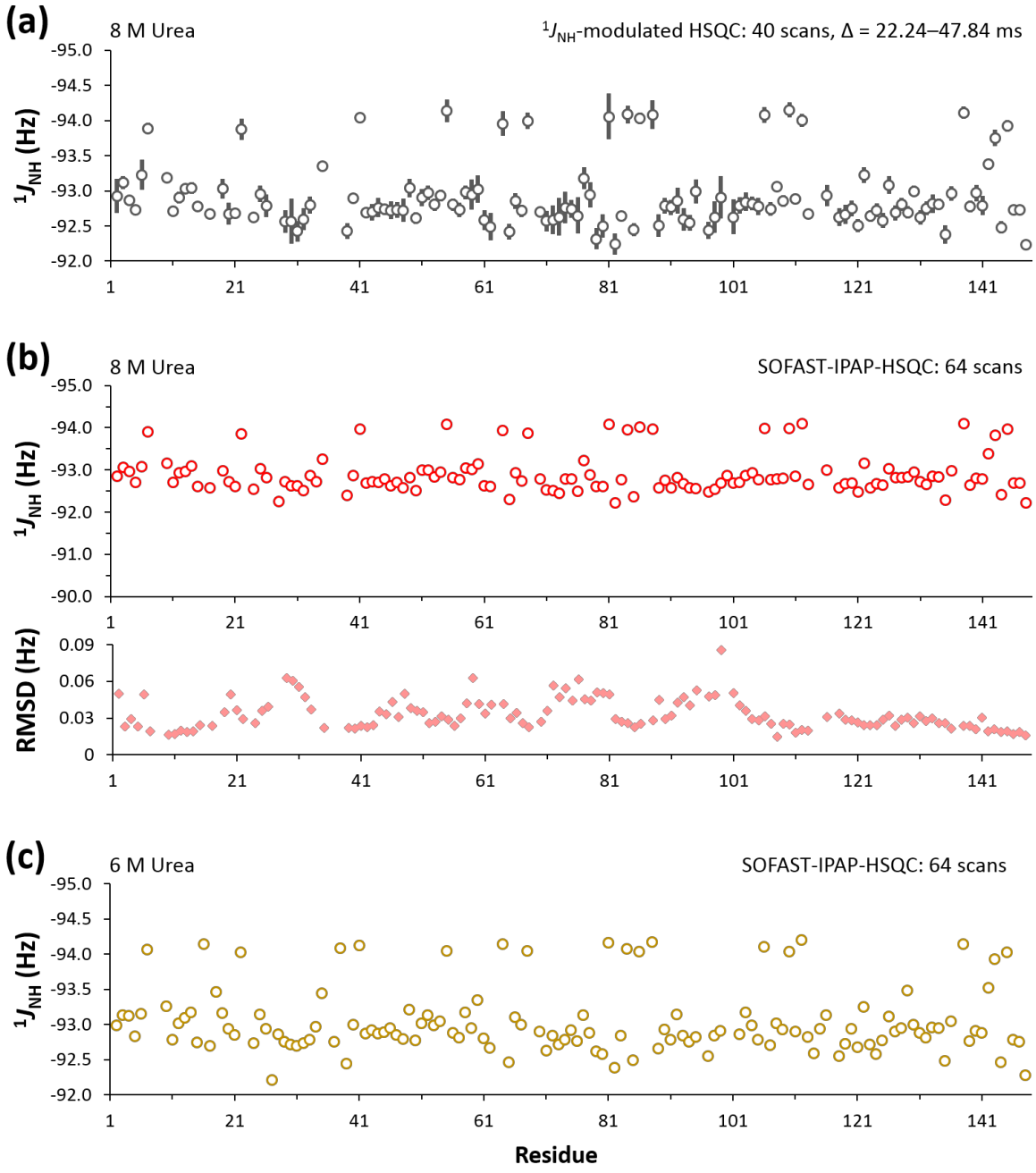
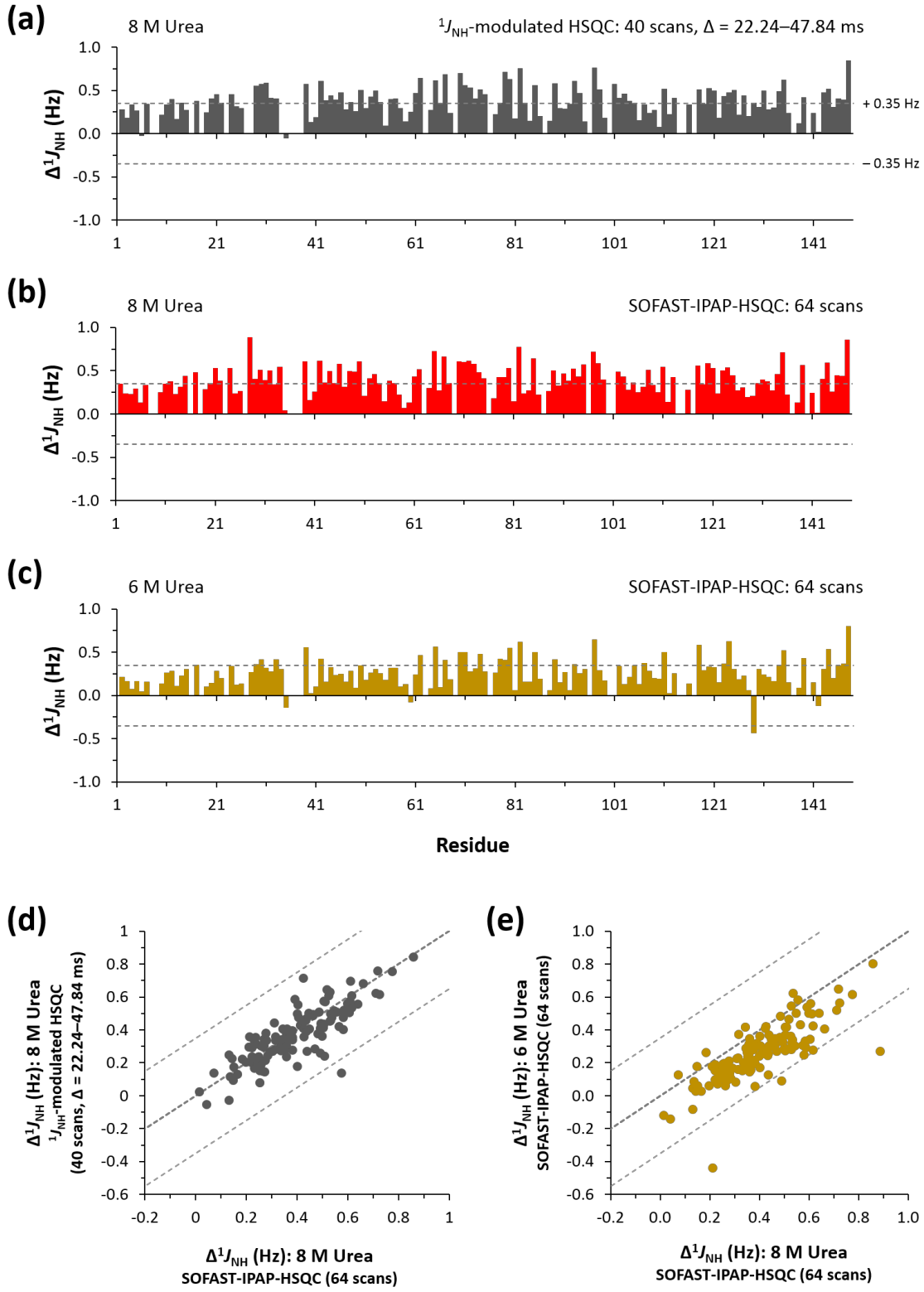


Figure 4.27: $^1J_{NH}$ values for OmpX in 8 M and 6 M urea
(a–b) $^1J_{NH}$ for OmpX in 8 M urea obtained by $^1J_{NH}$-modulated HSQC (40 scans, Δ = 22.24–47.84 ms) **(a)** and SOFAST-IPAP-HSQC (64 scans) **(b, top panel)**. The vertical lines in panel a show the range between the 2.5% and 97.5% confidence limits of the $^1J_{NH}$ values obtained from fitting of peak intensities with the sinusoidal decay function. The bottom half of panel b shows the RMSD of $^1J_{NH}$, calculated as the sum of RMSD for peak position for each component of the ^{15}N doublet. Residues without RMSD are due to missing information on the transverse relaxation rate derived from the corresponding $^1J_{NH}$-modulated HSQC (no peak detected or inadequate fitting of intensity modulation). **(c)** $^1J_{NH}$ for OmpX in 6 M urea obtained by SOFAST-IPAP-HSQC (64 scans). RMSD was not calculated because the corresponding $^1J_{NH}$-modulated HSQC was not recorded. All experiments were recorded at 288 K on a 800 MHz spectrometer.


(a)
8 M Urea
$^1J_{NH}$-modulated HSQC: 40 scans, Δ = 22.24–47.84 ms
Δ^1J_{NH} (Hz)
+ 0.35 Hz
– 0.35 Hz
(b)
8 M Urea
SOFAST-IPAP-HSQC: 64 scans
(c)
6 M Urea
SOFAST-IPAP-HSQC: 64 scans
Residue
(d)
Δ^1J_{NH} (Hz): 8 M Urea
$^1J_{NH}$-modulated HSQC
(40 scans, Δ = 22.24–47.84 ms)
Δ^1J_{NH} (Hz): 8 M Urea
SOFAST-IPAP-HSQC (64 scans)
(e)
Δ^1J_{NH} (Hz): 6 M Urea
SOFAST-IPAP-HSQC (64 scans)
Δ^1J_{NH} (Hz): 8 M Urea
SOFAST-IPAP-HSQC (64 scans)

(*Preceding page*)
Figure 4.28: Δ^1J_{NH} values for OmpX in 8 M and 6 M urea
(a–c) Δ^1J_{NH} are plotted against residue number. The horizontal grey dotted lines indicate the ± 0.35 Hz error range for the amino acid-specific random coil values reported by Xiang *et al* [201]. Several residues did not yield reliable data due to lack of assignments or were excluded from analysis due to peak overlap or there were < 13 reliable peak intensities across different Δ periods. **(d–e)** Δ^1J_{NH} values of OmpX in 8 M urea from SOFAST-IPAP-HSQC (64 scans) are compared against those of OmpX in 8 M urea from $^1J_{NH}$-modulated HSQC (40 scans) **(d)**, and OmpX in 6 M urea from SOFAST-IPAP-HSQC (64 scans) **(e)**. The expected linear correlation with gradient = 1 is shown as bold grey dotted lines. The light grey dotted lines represent the ± 0.35 Hz error range reported by Xiang *et al* [201].

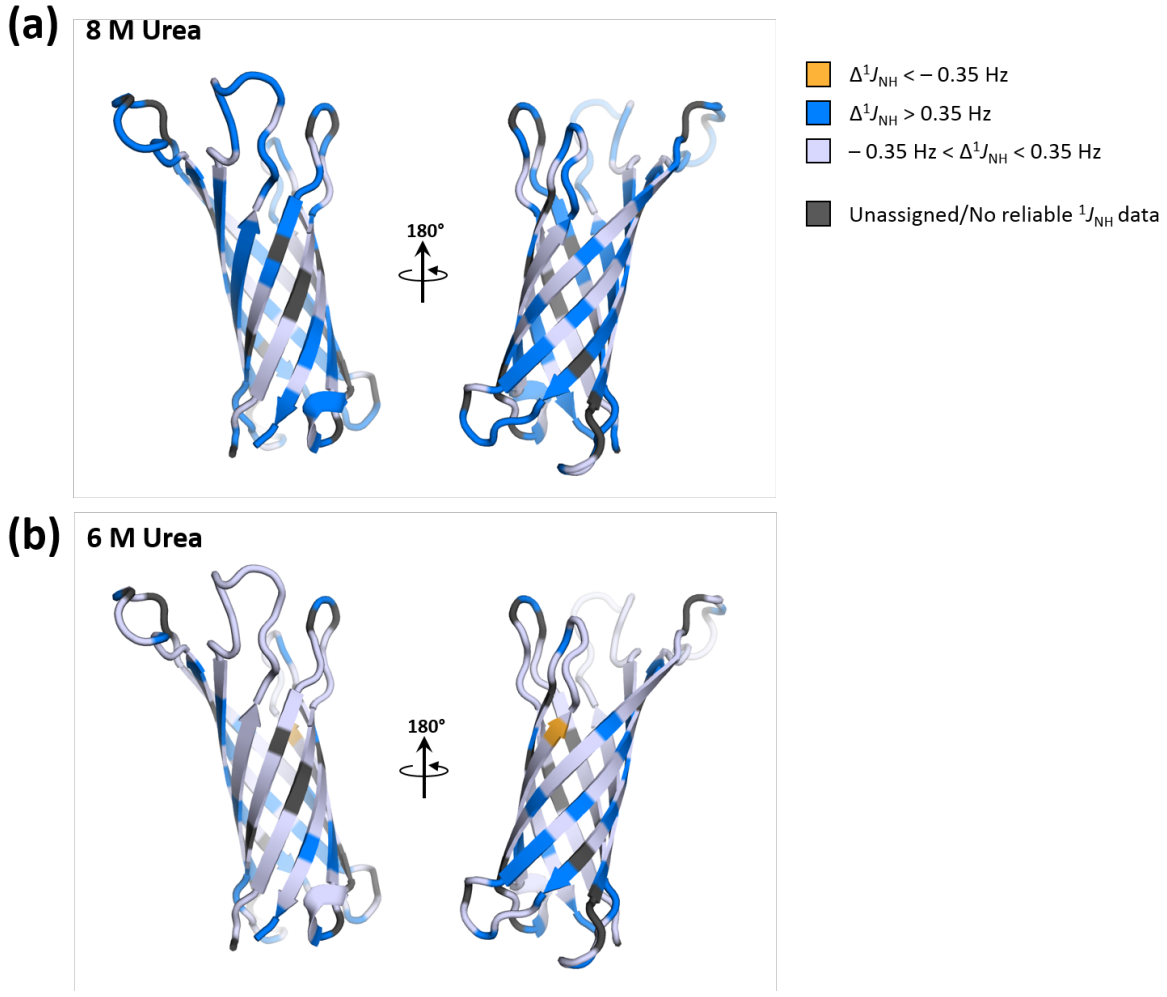


Figure 4.29: Δ^1J_{NH} values for OmpX in 8 M and 6 M urea from SOFAST-IPAP-HSQC
Δ^1J_{NH} are mapped onto the solution-state NMR structure of OmpX (PDB 2M07).

4.4 Discussion and conclusions

This chapter has explored the feasibility of studying H-bond formation during protein folding by measuring $^1J_{NH}$ using solution-state NMR. Proof-of-concept studies for extending the molecular weight limit of $^1J_{NH}$ measurements were first performed on the 20.9 kDa RalAΔC-GDP. Different implementations of IPAP-HSQC and $^1J_{NH}$-modulated HSQC were explored to study how the accuracy and precision of $^1J_{NH}$ were affected by the signal-to-noise ratio (number of scans), differential relaxation rates of in-phase and antiphase magnetisation (altered by the duration of the *J* evolution period) and lineshape (TROSY scheme). $^1J_{NH}$ measurements by IPAP-HSQC gave good correlations with H-bonds predicted by WhatIf but are plagued by large experimental errors. $^1J_{NH}$-modulated HSQC gave small experimental deviations but yielded big systematic errors. The large errors in both experiments arose from the utilisation of a protonated (rather than deuterated) protein sample. IPAP-HSQC and $^1J_{NH}$-modulated HSQC were subsequently attempted with OmpX-DPC, but both methods yielded poor correlation between experimentally-determined $^1J_{NH}$ and H-bonds predicted by WhatIf. $^1J_{NH}$ of OmpX in urea clearly indicate the lack of stable H-bonds, and no systematic deviation in $^1J_{NH}$ measured by $^1J_{NH}$-modulated HSQC.

In the IPAP-HSQC experiments, the predominant source of error is inaccuracies in determining the peak positions of the broad anti-TROSY lines (CSA(^{15}N)/DD$_{NH}$ CCR), leading to inaccurate $^1J_{NH}$ measurements from resonance splittings. Although this problem would also have been present in previous studies on ubiquitin (8.5 kDa; [201]) using the same NMR pulse scheme, the much larger sizes of RalAΔC-GDP (20.9 kDa) and OmpX-DPC (39.2 kDa; [285]) complex appears to have exacerbated the problem due to the substantially increased linewidth of the ^{15}N-anti-TROSY line, leading to much larger experimental errors.

A positive offset of 0.2–0.4 Hz was observed in Δ^1J_{NH} values derived from $^1J_{NH}$-modulated HSQC. This positive offset is consistent with and bigger than (as expected) the reported ~0.1 Hz deviation in $^1J_{NH}$ values of ^{15}N-labelled ubiquitin compared to those of [^{15}N,^{2}H]-enriched ubiquitin, and residues containing two vs. one β proton show greater discrepancy [367]. This discrepancy was not due to magnetisation evolution with CCR during

the long selective pulse, because similar $^1J_{NH}$ were obtained for non-deuterated and deuterated ubiquitin at different magnetic fields, in which the length of the selective pulse for achieving the same inversion bandwidth was substantially different [367]. This would indicate that the discrepancy in $^1J_{NH}$ values observed for OmpX are likely to also be due to incomplete refocusing of certain CCR pathways or more complicated relaxation mechanisms, possibly attributable to "forbidden" magnetisation transfer pathways and concomitant change in peak intensities and periodicity of the intensity modulation. Closer examination of the pulse sequence suggests that CSA(^{15}N)/DD$_{NR}$ was active in $^1J_{NH}$-modulated HSQC (and variations thereof), and could be the major contributor towards the systematic error (Figure 4.30). A possible solution is to apply one aliphatic proton-selective inversion pulse on either side of the amide proton-selective pulse during the *J* evolution period in the $^1J_{NH}$-modulated HSQC.

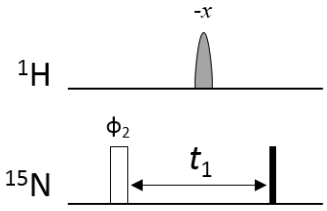
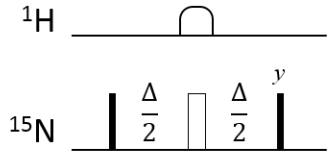
	IPAP-HSQC **SOFAST-IPAP-HSQC**	**$^1J_{NH}$-modulated HSQC** **$^1J_{NH}$-modulated TROSY-HSQC**
		
DD$_{NH}$/DD$_{NR}$		
CSA(^{15}N)/DD$_{NH}$	Evolves	
CSA(^{15}N)/DD$_{NR}$		Evolves

Figure 4.30: Active cross-correlated relaxation mechanisms in IPAP-HSQC and $^1J_{NH}$-modulated HSQC implementations
The *J* coupling evolution periods of the different experiments are extracted from Figures 4.2 and 4.4. CSA(^{15}N)/DD$_{NH}$ is active in IPAP-HSQC experiments. CSA(^{15}N)/DD$_{NR}$ is active in $^1J_{NH}$-modulated HSQC experiments. Other CCR mechanisms are, in principle, refocused.

Differential relaxation rates of in-phase and antiphase magnetisation also contributes towards inaccuracies in $^1J_{NH}$-modulated HSQC. Although the final signal detected in the ^{1}H channel originates from the refocused in-phase operator after the *J* evolution period, anti-phase magnetisation is still generated at various points in the pulse sequence. The larger relaxation

rate constants of anti-phase magnetisation contribute towards faster signal intensity decay and an apparent leftward-shift of $^{1}J_{NH}$-modulated intensity curves, hence giving an apparent positive deviation for most $\Delta^{1}J_{NH}$ of RalAΔC-GDP and OmpX-DPC measured by $^{1}J_{NH}$-modulated HSQC.

Interestingly, the 0.2–0.4 Hz offset in $\Delta^{1}J_{NH}$ of the 20.9 kDa RalAΔC-GDP measured by $^{1}J_{NH}$-modulated HSQC were not observed in comparable measurements performed with OmpX in 8 M urea. This likely reflects that in denatured OmpX, (1) there is lack of correlation between rotational correlational time and molecular weight, and (2) CCR is reduced due to loosened restrictions on dihedral angles, in turn, less favourable internuclear distances and geometry in $^{15}N\text{-}H_{\alpha/\beta}$ dipoles. These observations can be supported by the small discrepancy in $^{1}J_{NH}$ of glycines and flexible residues reported for non-deuterated vs. deuterated ubiquitin [367].

It was initially conceived that method optimisation would enable measurement of $^{1}J_{NH}$ in folded and denatured pSRII. This would have provided a good foundation for complementary studies on identifying the donor and acceptor atoms of individual H-bonds and characterising H-bond geometries by measuring $^{3h}J_{NC'}$ couplings. Such studies have not been pursued, since the results from this chapter indicate that very big experimental errors will be expected for protonated samples of pSRII in c7-DHPC (~70 kDa; [216]). Perdeuterated membrane protein samples, preferably reconstituted in deuterated detergents, must be utilised for future $^{1}J_{NH}$ studies. The methodology for optimising NMR pulse sequences in this chapter would provide a solid foundation for reassessing the effects of different parameters on the precision and accuracy of $^{1}J_{NH}$ measurements. Depending on the sensitivity of $^{1}J_{NH}$ measurements on perdeuterated samples, future studies could include correlation of experimentally-determined $^{1}J_{NH}$ with different H-bond or protein backbone geometries from solution-state NMR structures, and other *J*-coupling constants. Future studies could also include determining the random coil $^{1}J_{NH}$ value for tryptophan, which would be useful considering their abundance in membrane proteins.

5 Towards assembling pSRII using split intein

5.1 Introduction

Studies on protein assembly from complementary fragments have underpinned the two-stage model and the long-range interactions model for α-helical membrane protein folding. The native chromophore and proton-pumping function of bacteriorhodopsin can be successfully reconstituted by assembling complementary combinations of proteolytic fragments [124,125] and/or complementary synthetic peptides [126,127], or complementary or overlapping fragments which have been individually expressed in and purified from *E. coli* [128]. Absence of any transmembrane segments results in abolishment of pigment formation. These results suggest that the transmembrane helices are indispensable in the folding of bacteriorhodopsin while the loop regions are dispensable, despite having some contribution towards the stability of folded bacteriorhodopsin [128]. On the other hand, refolding of rhodopsin from helical fragments has only been shown using fragments which are separated in the second or the third cytoplasmic loops, indicating the indispensability of the remaining loops in the folding process [136]. Extracellular loops are important for the folding of rhodopsin. *In vitro* folding studies [135,137,138] suggest that interactions between the extracellular and transmembrane domains constitute a "folding core" which is important in the early stages of folding. Moreover, most point mutations [382] which cause Retinitis Pigmentosa, a retinal degenerative disease due to misfolding of rhodopsin [383], are located in the extracellular loops [139].

The aim of this chapter is to establish a method for assembling pSRII from complementary fragments *in vivo*, with the vision of applying said method towards studying the roles of individual loops and helices in the folding pathway of pSRII.

Inteins are proteins encoded within a precursor, and are capable of excising themselves from the precursor after translation and ligating the flanking amino acid sequences with a peptide bond. This process is called protein *trans*-splicing [384]. In their native biological contexts,

inteins are embedded into the open reading frame of host proteins. The host protein sequences which flank the intein are called exteins. Full-length inteins contain the N- and C-terminal splicing domains which flank a homing endonuclease domain that is not required for splicing. Mini inteins lack the endonuclease domain [385,386].

Split inteins can be naturally-occurring or generated artificially by splitting contiguous inteins. They consist of separately-encoded N- and C-terminal fragments (Int^N and Int^C), which first assemble and are then excised, and the N- and C-exteins are joined together. Naturally-occurring split inteins generally have a relatively long Int^N (~100–120 residues) and a short Int^C (~35 residues) [386]. The cyanobacterial DnaE intein family is the largest known class of split inteins. In their biological contexts, DnaE split inteins assemble the catalytic subunit of DNA polymerase III (DnaE). Several members of this family, including DnaE from *Synechocystis* sp. strain PCC6803 (*Ssp*) and *Nostoc punctiforme* (*Npu*), have been well-characterised, and carry out protein *trans*-splicing with high efficiency, thus making them ideal tools for protein engineering [387].

Split intein fragments interact through a multiphase "capture and collapse" process [388]. The *Npu* DnaE Int^C is highly flexible, based on NMR relaxation experiments and poor dispersion in the [^{1}H,^{15}N] HSQC. Int^N is comprised of two structurally distinct lobes, formed by residues 1–50 (Int^{N1}) and 51–102 (Int^{N2}), respectively. The initial association ("capture") between Int^N and Int^C is mediated by electrostatic interactions of the disordered and highly basic (pI = 9.7) Int^C with the extended and highly acidic Int^{N2}. The two disordered regions are hypothesised to provide larger capture radii for enhanced probability for association. The Int^{N2}-Int^C complex is an intermediate with secondary structure and intermolecular interactions resembling half of the native intein complex. The N- and C-terminal fragments associate with low nanomolar affinity for both *Ssp* [389] and *Npu* [390] split inteins. Next, the flexible regions "collapse" into an ordered state, as evidenced by the more compact nature of Int^{N2} upon binding to Int^C as determined by SEC-MALS. The intermediate is further stabilised by hydrophobic interactions between Int^C and the well-ordered Int^{N1}, as Int^C intercalates between Int^{N1} and Int^{N2} to form an interwoven topology. Effectively, the disordered regions of the precursors provide enhanced probability for association and access to an interwoven topology *in trans*, while the combination of electrostatic and hydrophobic

interactions in the Int^N-Int^C complex contributes towards the fast on-rate and slow off-rate of the complex. Formation of a stable complex is crucial for the subsequent multistep splicing reaction.

A self-catalysed mechanism was proposed based on mutagenesis and chemical cleavage studies on an *in vitro* splicing system which contains the contiguous intein from the DNA polymerase of the archaeon *Pyrococcus species* GB-D (Psp pol intein-1) [391]. In the study, mutations were used to block various steps in the protein splicing pathway, allowing each step to be studied independently by isolating and analysing the splicing precursors, intermediates and products.

Upon association between Int^N and Int^C, a conserved Cys or Ser at the N-terminus of Int^N initiates a nucleophilic attack on the carbonyl carbon of the C-terminal N-extein residue [384] (Figure 5.1). Next, an N-to-O/S acyl shift leads to the transfer of the N-extein to the side-chain of the first residue of the C-extein (the +1 residue), forming a linear (thio)ester intermediate. This acyl shift is mediated by a conserved threonine and histidine in the TxxH motif of Int^N, along with the aspartate in the DHNF motif in Int^C [392]. The nucleophilic +1 residue of the C-extein attacks the (thio)ester bond of the linear intermediate, resulting in the formation of a branched ester intermediate, where the N-extein is linked to the C-extein by an (thio)ester bond. The last residue of Int^C, typically Asn, forms a cyclic structure, whereby Int^C is cleaved off from the branched intermediate. His-125 in the DHNF motif acts as the general acid/base in the resolution of the branched intermediate, and is completely conserved in the DnaE family [393,394]. Asn cyclisation is the rate-limiting step for *trans*-splicing by *Ssp* DnaE [395], whereas all steps occur with similar rates for *Npu* DnaE [394]. The last step is a spontaneous S–N acyl rearrangement between the (thio)esterified exteins because peptide formation is more energetically favourable, resulting in the final formation of the peptide bond and yielding the ligated host protein [384]. After *trans*-splicing, the three flanking amino acids at each end of the split intein are included in the ligated product, and are referred to as the intein "scar" [396].

There are several variations of the canonical *trans*-splicing mechanism [392]. Inteins which lack the nucleophilic Cys or Ser in Int^N (e.g. KlbA family of inteins) cannot initiate splicing

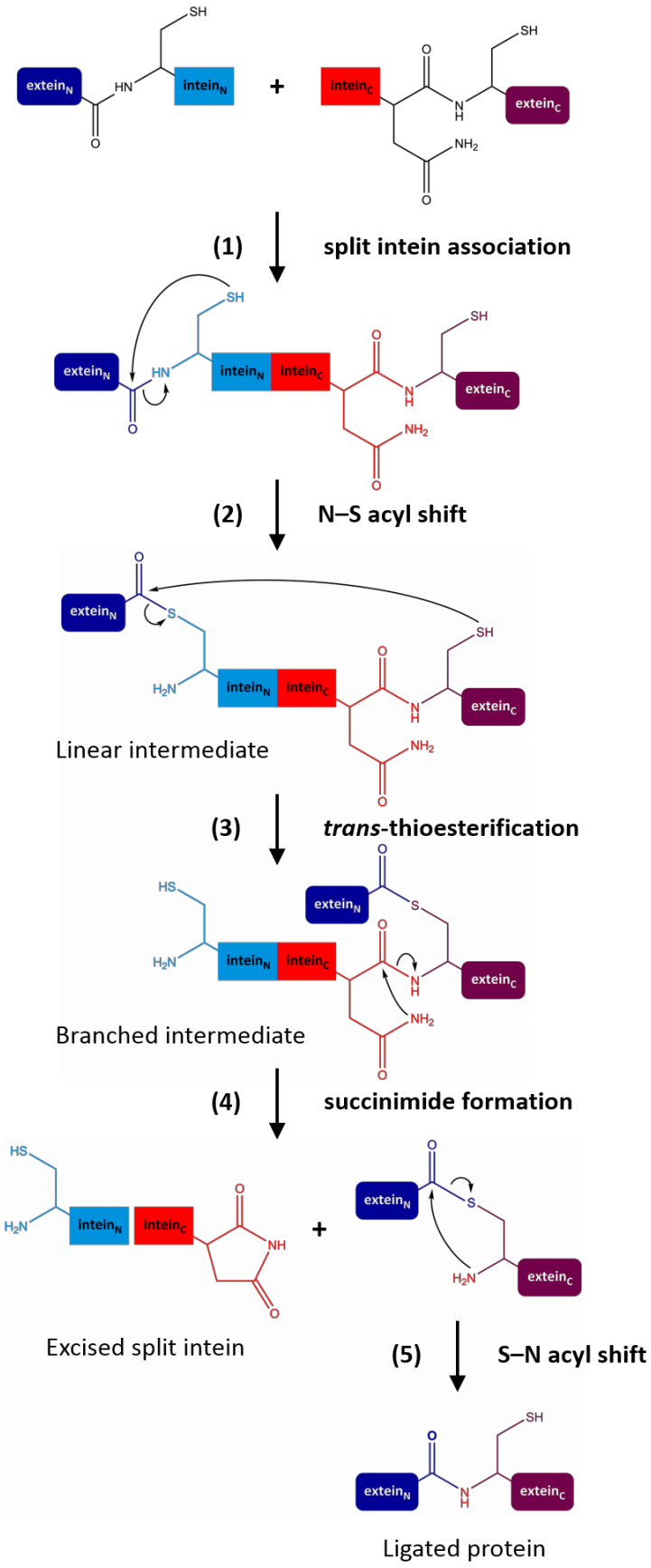
through formation of the linear (thio)ester intermediate. These inteins could directly form the branched intermediate upon N-terminal activation given that the N-terminal peptide bond of Int^N is destabilised or twisted, making it more susceptible to nucleophilic attack [397]. Alternatively, a different branched thioester intermediate could be formed using a unique Cys in Int^C [398]. A few inteins also have Glu or Asp rather than Asn as the last residue of Int^C. The reaction would theoretically proceed through the same chemical mechanism [392], albeit with the formation of a different cyclic by-product and with different efficiency – molecular dynamics simulations suggest that an Asn side-chain samples the conformation required for nucleophilic attack more frequently than a Gln side-chain [399].

There are two common side reactions which compete against protein *trans*-splicing [391] (Figure 5.2). N-extein cleavage can occur when an external nucleophile, commonly water or a thiol, attacks the linear or branched (thio)ester intermediates to yield free N-extein. C-extein cleavage can occur when the C-terminal Asn cyclises prior to formation of the branched intermediate, releasing the C-extein. Mutation of critical intein or extein residues, or alteration of experimental conditions, can favour cleavage over splicing for protein engineering purposes.

(*next page*)

Figure 5.1: Canonical protein *trans*-splicing mechanism utilised by split inteins

The N-precursor ($extein_N$–$intein_N$ fusion) and C-precursor ($extein_C$–$intein_C$ fusion) proteins are expressed from separate plasmids. Upon association of $intein_N$ and $intein_C$ (step 1), an N–S acyl shift involving the first residue of $intein_N$ occurs (step 2). The linear thioester intermediate is attacked by the first residue of $extein_C$ (step 3), resulting in *trans*-thioesterification reaction and formation of a branched thioester intermediate. At this stage, $intein_N$ is physically separated from $extein_N$. The last residue of $intein_C$ (usually Asn) circularises to form a succinimide ring (step 4), effectively cleaving $intein_C$ from the thioesterified exteins. A spontaneous S–N acyl shift between the esterified exteins leads to peptide bond formation between $extein_N$ and $extein_C$ (step 5). Figure adapted from [384].


extein$_N$
intein$_N$
intein$_C$
extein$_C$
(1) split intein association
(2) N–S acyl shift
Linear intermediate
(3) *trans*-thioesterification
Branched intermediate
(4) succinimide formation
Excised split intein
(5) S–N acyl shift
Ligated protein

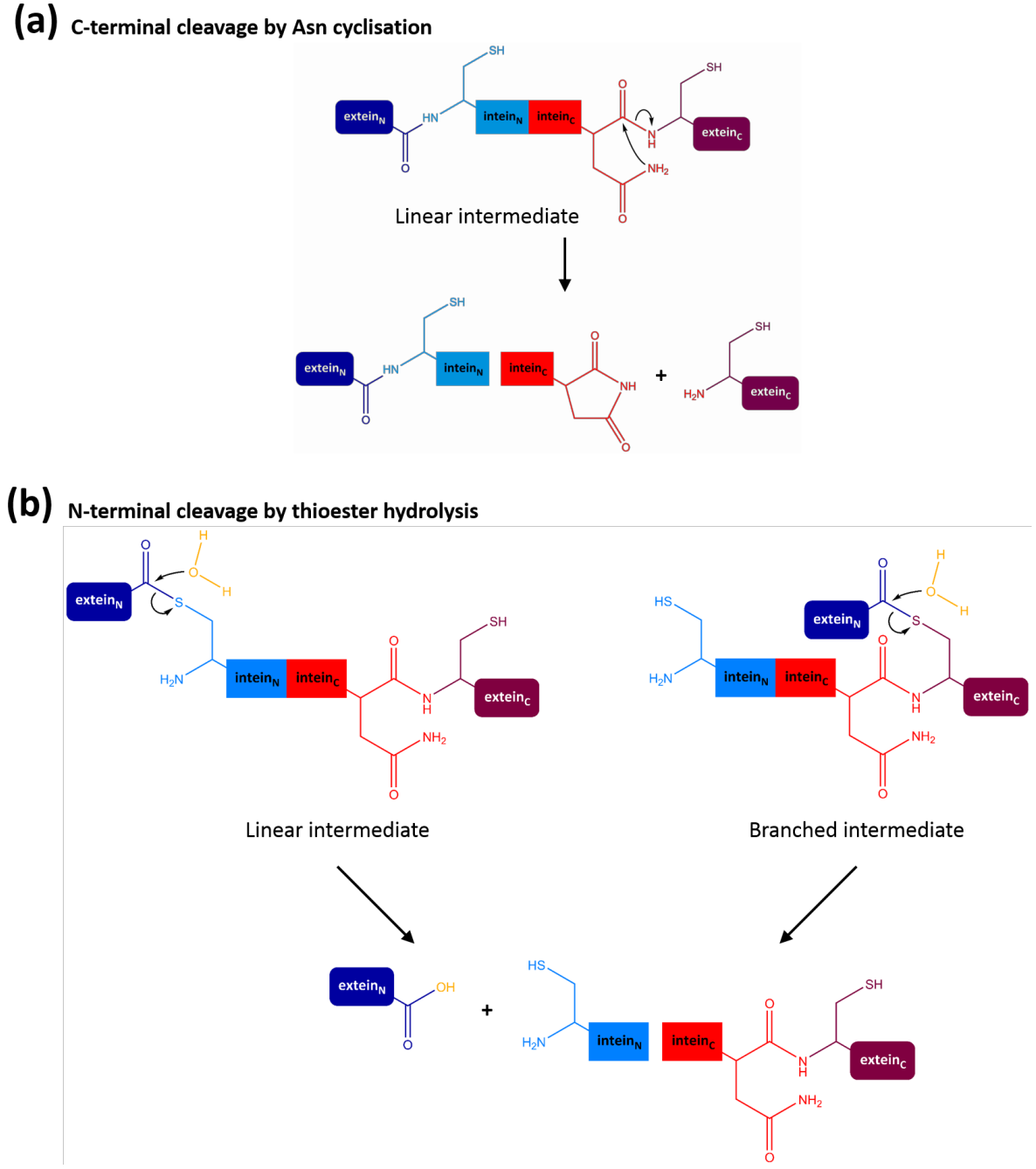


Figure 5.2: Mechanisms for N-terminal and C-terminal cleavage side reactions
(a) C-terminal cleavage results from cyclisation of Asn (last residue of intein$_C$) prior to formation of the branched intermediate. Asn cyclisation can also occur prior to or after N-terminal cleavage. **(b)** N-terminal cleavage results from ester bond hydrolysis in the linear or branched thioester intermediates. Figure adapted from [391].

Inteins are generally tolerant to a wide range of extein sequences, and can therefore be applied towards *trans*-splicing of different non-native protein fragments. Several extein residues which are proximal to the intein sequence can greatly influence the efficiency of the splicing reaction. The native splice site of *Npu* DnaE is Ala(–3)-Glu(–2)-Tyr(–1)/Cys(+1)-Phe(+2)-Asn(+3), where residues with negative numbers are from the N-extein, residues with positive numbers are from the C-extein and the slash represents the splice junction. These six residues constitute the scar upon splicing. A study on different combinations of –3 to +3 extein residues with the *Npu* DnaE intein has been done using genetic selection for kanamycin resistance conferred by successful assembly of *C. diphtheriae* aminoglycoside phosphotransferase (*Aph*) [396]. The range of N-terminal extein sequences is very broad, with the importance of flexibility being noted from the abundance of Gly hits at the –3 and –2 positions. This is consistent with the importance of the disordered nature of *Npu* DnaE Int^{N2} in mediating its association with Int^{C}. Different amino acids, including Ser, Asn, Tyr, Phe, His, Asp, Glu, Lys and Gln, can be accommodated as the –1 residue with little influence on the splicing activity [400]. The +1 residue (the first residue in the C-extein) must contain a thiol or hydroxyl group (i.e. Cys, Ser or Thr) for successful formation of the linear and branched (thio)ester intermediates. Large specificity preferences have been observed at the +2 and +3 positions. Sequences containing Trp and Met at the +2 position can splice as efficiently as the native extein. The crystal structure of the *Ssp* DnaE intein (PDB 1ZDE) shows close packing of the native Phe(+2) against His-125 of the DHNF motif in Int^{C}, thereby stabilising His-125. MD simulations show that mutations at the +2 position alter the dynamics and side-chain rotamer conformations of His-125, and positioning of the C-intein/C-extein junction relative to the His-125 loop. These changes affect the rate at which the branched intermediate can be resolved [394]. There is a preference at the +3 position for residues similar to the natural Asn, including Tyr, Asp, Gln and Glu [396]. Hence, *trans*-splicing using the *Npu* DnaE split intein can be performed for a range of extein sequences, suggesting the potential applicability of this split intein system with minimal/no introduction of scar sequences.

Npu DnaE intein shows robust *trans*-splicing activity with an efficiency of >98% such that little precursor proteins remain, and is more efficient than that of *Ssp* DnaE. *Npu* DnaE also shows greater tolerance to extein sequence variations, hence widening its application [387].

Npu DnaE with native flanking extein sequences has fast *trans*-splicing kinetics, with an apparent rate constant (K) of 0.63 ± 0.07 min^{-1} and a $t_{1/2}$ of 66 s at 30 °C. Several non-native extein sequences display comparable or faster *trans*-splicing rates relative to the native extein sequence. For example, the extein sequence used for studies in this chapter, Ala-Glu-Lys/Cys-Trp-Leu is reported to have $K = 0.9 \pm 0.3$ min^{-1} and $t_{1/2} = 46$ s at 30 °C [396].

Protein *trans*-splicing can be performed both *in vitro* and *in vivo* because it is an auto-catalytic process which requires only the formation of active split intein for protein ligation. There are two approaches for *in vitro* protein *trans*-splicing. Precursor fragments can be purified in their denatured forms and refolded *in vitro* to restore protein *trans*-splicing activity. Alternatively, native chemical ligation involves ligation of a soluble precursor containing an α-thioester with the other precursor bearing an N-terminal cysteine. Both methods require separate preparation of precursor fragments and extensive optimisation of ligation conditions.

In comparison, protein *trans*-splicing *in vivo* does not require individual preparation of precursor fragments nor any additional chemical reagents, and is therefore less labour-intensive. Given its convenience and versatility, *in vivo* protein *trans*-splicing has been used successfully in a wide range of applications [401], including segmental isotope labelling for NMR studies [400,402], ligation of synthetic molecules (e.g. fluorescent tags) to proteins (protein semi-synthesis) [403], selective chemical modification of proteins [404], and production of cyclic peptides and proteins [405].

The two precursors can be co-expressed or expressed sequentially (dual-expression) in *E. coli*. To function, the two split intein fragments must find one another efficiently and form a stable complex which can persist throughout the multistep protein *trans*-splicing reaction, which may vary from seconds to hours between different naturally and artificially split inteins [395,401,406–408]. The first crucial step is to define an appropriate location to divide the protein-of-interest into two segments [400]. The split site is usually taken as a linker region connecting domains within the protein-of-interest to minimise structural disruption. Since a short sequence corresponding to the intein scar would be inserted and these proximal extein residues greatly influence the splicing yield, appropriate sequences near the splicing junctions

must be chosen for both the N- and C-exteins to ensure high splicing efficiency and rate whilst minimising disruption to the native structure of the protein-of-interest.

The appropriate cloning strategy must then be applied in order to obtain the appropriate constructs. The expression system developed by the Iwai lab is a dual-expression system designed for *in vivo* segmental isotopic labelling [400]. The N- and C-terminal precursors are cloned into two separate plasmids such that their expression are induced by different inducers. It is important to optimise expression conditions to obtain high yields of the N- and C-terminal precursors. Protein *trans*-splicing must be tested in small-scale expression tests, in which the timing of inductions, growth temperatures and concentrations of inducers must be optimised for maximising ligation products in SDS-PAGE analysis [400]. While splicing of native host proteins is very rapid and precursor proteins are rarely detected in nature, splicing of heterologous host proteins often results in unreacted precursor or splice junction cleavage by-products, thus yielding multiple bands on SDS-PAGE.

The solubility of precursor proteins fused with split intein fragments could be a limiting factor. At least one of the two precursors must be soluble for successful *trans*-splicing. Previous studies have assumed that protein *trans*-splicing might be more efficient in the cytosol since the intein segments would have higher degrees of freedom for achieving the required orientation to associate with each other. *Trans*-splicing of membrane proteins is particularly challenging. The precursors must remain stable and soluble in the cytosol until protein *trans*-splicing occurs, and the ligated protein must then translocate and insert into the membrane fraction to ensure functional folding [228]. It remains unclear whether protein *trans*-splicing could occur once the precursors have localised to the membrane fraction. The signal peptide sequence can be omitted to inhibit fast membrane insertion of the N-precursor. Solubility enhancement tags, such as yeast Smt3 protein [409], glutathione S-transferase (GST), B1 domain of the immunoglobulin-binding protein G (GB1) [410] and maltose-binding protein (MBP) [411], can be attached to enhance expression yield and solubility of each fragment. These tags should ideally be removed or produced in the unlabelled form for NMR studies to avoid spectral crowding. Potential structural perturbation of the fusion protein due to the presence of these tags must also be carefully evaluated [400].

Trans-splicing could also continue during protein purification under conditions which permit co-purification of the unspliced precursor proteins [402]. The *Npu* DnaE split intein is highly robust, and has been shown to give at least 65% yield at temperatures from 6 to 37 °C and in the presence of up to 6 M urea [401]. Optimisation of reaction conditions for longer incubation of precursor proteins *in vitro* could also help improve the yield of ligated product.

In this chapter, assembly of functional pSRII was attempted in *E. coli* using two complementary fragments of pSRII which are fused with the N- and C-terminal *Npu* DnaE split intein. Successful production of ligated pSRII with covalently-bound retinal chromophore absorbing at the correct wavelength would indicate that the splice site has minimal effects on pSRII folding. The ligated pSRII should purify in a similar manner as WT-pSRII, hence circumventing the need for extensive optimisation of the purification protocol. With suitable control experiments and method for quantifying the yield of ligated pSRII, protein *trans*-splicing using split intein can be further applied towards studying the roles of individual loops and helices in the folding of pSRII *in vivo*.

5.2 Results

5.2.1 Construct design

Design of precursor constructs involve selecting a split site in pSRII and appropriate residues which flank the intein fragments, both of which are critical in ensuring successful and efficient protein *trans*-splicing. Ideally, there should be minimal/no differences between the ligated product and full-length pSRII mutant which accounts for insertion of intein ligation scar residues.

The loop regions in pSRII were assessed to identify a suitable splice site. Microbial rhodopsins have three extracellular and three intracellular loops. Several loops are known to have important roles in the structure and function of pSRII. For example, the interaction of pSRII with its cognate transducer, pHtrII, is directly mediated by the EF loop [412,413] and also leads to chemical shift changes in the FG loop of pSRII [278]. A study on chimeras of pSRII and bacteriorhodopsin shows important roles of the DE loop in the positioning of helix E and, in turn, colour tuning of the two proteins [414]. Splitting pSRII in loop regions with known involvement in the structure and function of pSRII could adversely affect protein stability, and are not ideal candidates for proof-of-concept studies on the application of split inteins. The stability of individual helices or helical fragments must also be considered. While this information is unavailable for pSRII, circular dichroism of individual fragments of bacteriorhodopsin in DMPC/DHPC bicelles shows that the C-terminal CD, CE, DE, DG, EG and FG fragments do not form helices of the expected lengths, with the FG fragment displaying high amount of non-native β-sheet conformation [415]. Splitting pSRII at N-terminal loops might be more preferable, although it must be noted that the stability of helical fragments depends on amino acid sequence and the membrane environment.

The BC loop was chosen as the splice site in this study. The BC loop is usually the longest loop and its conformation is varied the most amongst different microbial rhodopsins. Superposition of the C_α traces of pSRII, bacteriorhodopsin and halorhodopsin shows marked differences in the BC loops of the three proteins [336]. The BC loops of pSRII and bacteriorhodopsin [214] are organised as β-hairpins and cap the extracellular side of the proteins. In pSRII, the hairpin is partially sealed by hydrogen bonds between Arg-66 and

residues at the end of helix D (Figure 5.3b). The corresponding Arg-82 in *Haloquadratum walsbyi* bacteriorhodopsin shows unique a H-bonding network with Thr-201, which links the BC and FG loops to shield the retinal-binding pocket from the extracellular environment [416]. Amino acid sequence alignment shows that the BC loop of pSRII is one of the shortest amongst those of microbial rhodopsins from different halobacteria [417]. This might suggest that the BC loop of pSRII could be tolerant towards extra residues being inserted as the intein ligation scar. The sequences of the *Npu* DnaE splice junction is Ala-Glu-Tyr/Cys-Phe-Asn. To avoid disruption of H-bonds between the two β-strands of the hairpin structure, the splice site was introduced in the linker region between the strands. This linker region consists of the residues Ala-64, Glu-65 and Arg-66 in WT-pSRII. To minimise the number of inserted residues, the construct encoding for the N-precursor was designed such that Ala-64 and Glu-65 of pSRII were positioned as the –3 and –2 residues of the N-extein. Only 2 out of the 21 combinations of extein sequences investigated by Cheriyan *et al*. contain Ala and Glu at the –3 and –2 positions: Ala-Glu-Tyr/Cys-Phe-Asn (the native extein sequence; abbreviated as AEY/CFN) and Ala-Glu-Lys/Cys-Trp-Leu (AEK/CWL). The latter was chosen in this investigation because (1) the Lys at the –1 position is chemically more similar to Arg-66 of WT-pSRII, and (2) the AEK/CWL extein sequence yields slightly faster protein *trans*-splicing kinetics [396]. The construct encoding for the C-precursor was designed to preserve Arg-66, considering its potential importance in anchoring the BC loop in place via H-bonds. In short, in the final ligated product, the residues KCWL would be inserted between the original Glu-65 and Arg-66, thus lengthening the linker region of the hairpin by four residues (Figure 5.3a).

Next, the appropriate affinity tags must be incorporated for the purposes of detecting expression yield, identifying ligation products and protein purification. The C-precursor and ligated pSRII-KCWL are both His_6-tagged and have similar molecular weights (25.7 kDa and 27.0 kDa, respectively). This indicates potential problems with resolving the two species on SDS-PAGE and Western blots probing only for the His_6 tag. A Strep-tag II was therefore added to the N-terminus of the C-precursor. Expression of the C-precursor yields a 25.7-kDa band which is detectable by probing for His_6 and Strep tags. With successful ligation, the C-intein is cleaved from the C-precursor along with the Strep tag. In principle, this would lead to decreased intensity of the 25.7-kDa band (detectable by both His_6 and Strep probes)

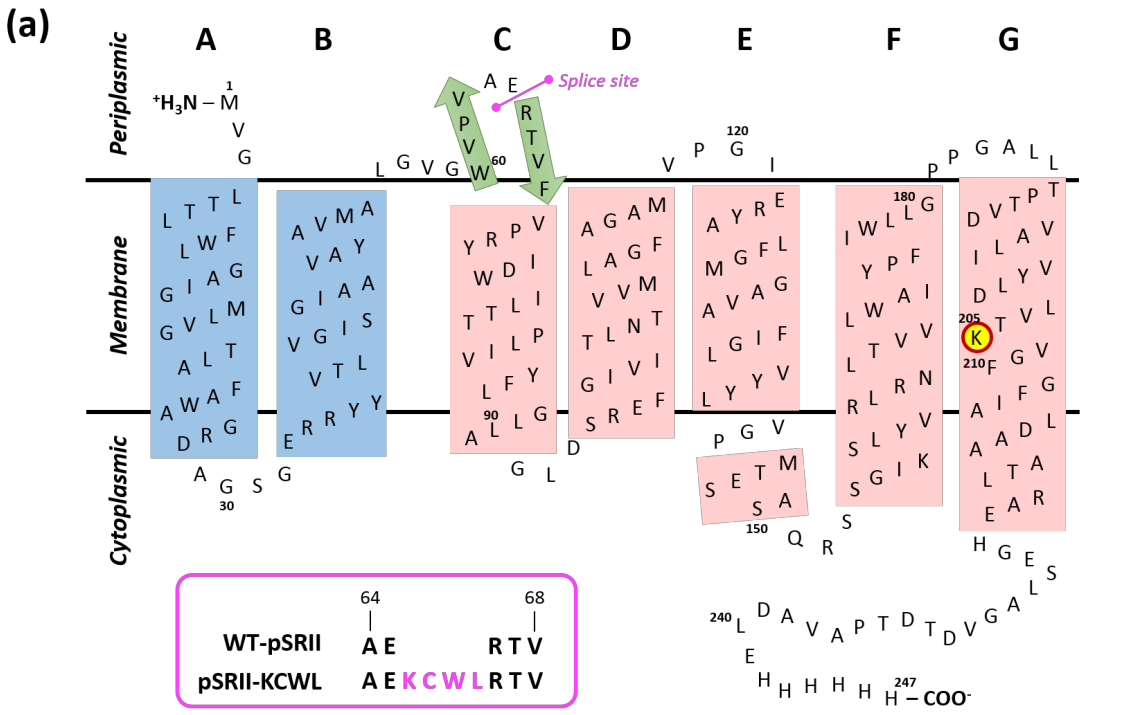
and the appearance of bands at 5.5 kDa and 13.1 kDa corresponding to the excised C-intein and N-intein, respectively, both of which are detectable only by Strep probe (Figure 5.3c).

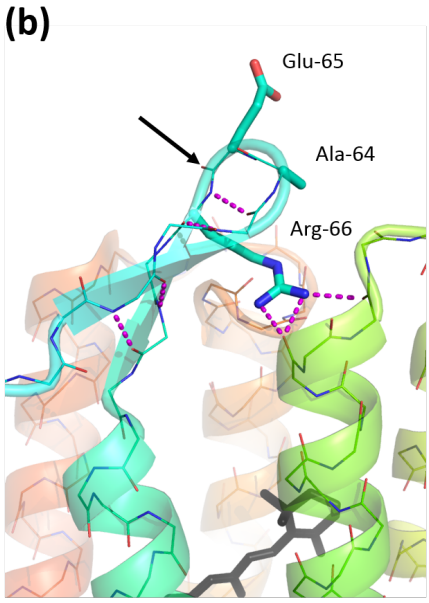
To summarise, constructs were designed to encode for pSRII^{N}(KCWL)-DnaE^{N}-Strep (the N-precursor) under the control of the T7 promoter in a pET28b(+) vector, and Strep-DnaE^{C}-pSRII^{C}(KCWL)-His$_6$ (the C-precursor) under the control of the arabinose-inducible PBAD promoter in a pBad vector. These constructs were produced by PCR and Gibson assembly using the dual-expression system made available by the Iwai lab [387] and WT-pSRII in pET28b(+) as templates (see Materials & Methods).

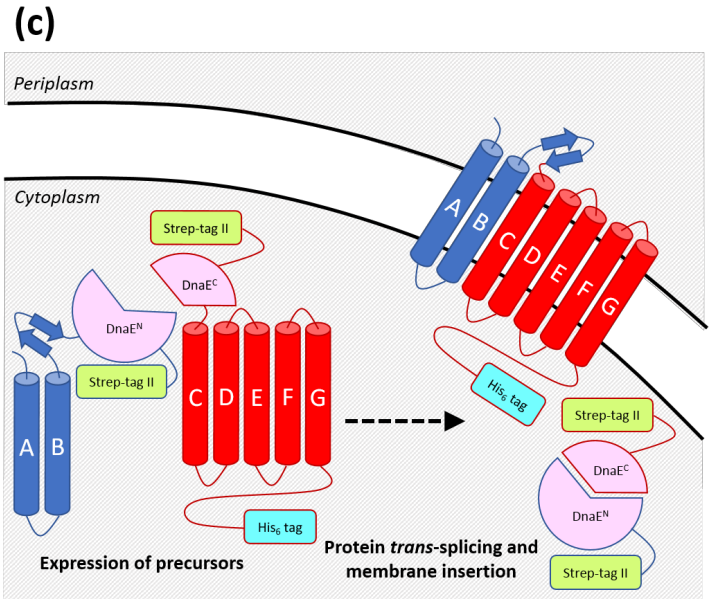
(*Next page*)

Figure 5.3: Constructs for *in vivo trans*-splicing of pSRII

(a) 2D representation of the secondary structure of pSRII. Lys-205, the site of retinal Schiff base attachment, is circled in red. To ensure efficient *trans*-splicing reaction by *Npu* DnaE split intein, the four residues Lys-Cys-Trp-Leu (KCWL) were inserted in the middle of the BC loop and constitute the splice junction. Helices which are part of the N-precursor are coloured blue, and those in the C-precursor are coloured red. **(b)** The BC loop of pSRII forms a β-hairpin structure. The side-chains of residues 64–66, which join the two β-strands of the hairpin, are shown using stick representation. The side-chain of Arg-66 forms H-bonds with the backbone carbonyls of Ala-116 and Val-118 at the end of helix D (PDB 1H68; [336]). This Arg residue was therefore preserved in the C-precursor to minimise structural changes. The splice site is indicated with an arrow. **(c)** Schematic diagram illustrating *in vivo* protein *trans*-splicing of two fragments of pSRII using *Npu* DnaE split intein. The N-precursor consists of helices A–B, the first β-strand of the hairpin constituting the BC loop, DnaE^{N} and a Strep tag. The C-precursor consists of a Strep tag, DnaE^{C}, the second β-strand in the BC loop, helices C–G and the C-terminal tail of pSRII followed by a His$_6$ tag. Upon successful *trans*-splicing, both Strep tags and the intein should be excised, yielding full-length pSRII(KCWL)-His$_6$.







5.2.2 Expression and purification of full-length pSRII-KCWL

To confirm the stability and retinal-binding function of the ligated protein with insertion of four non-native residues (Lys-Cys-Trp-Leu, or KCWL) in the BC loop as the intein scar, a full-length pSRII-KCWL mutant with the same BC loop was prepared. pSRII-KCWL in pET28b(+) vector was obtained by site-directed mutagenesis of WT-pSRII. The protein was expressed and purified in a similar manner as WT-pSRII, up to and including the 60-h

incubation at 55 °C to precipitate contaminating proteins (see Materials & Methods). The retinal chromophore of pSRII-KCWL had a similar absorption profile as WT-pSRII, with λ_{max} = 498 nm and visible shoulder peaks at 460 and 420 nm indicative of a tight retinal binding pocket.

Assuming the same extinction coefficient at 498 nm, pSRII-KCWL in DDM had an estimated purity of 69%, whereas WT-pSRII in DDM typically has 75–85% purity (78% in Figure 5.4b). Misfolded pSRII-KCWL absorbing at 390 nm was clearly observable. In an attempt to separate well-folded pSRII-KCWL (λ_{max} = 498 nm) from the misfolded species, the sample was subjected to size exclusion chromatography (Figure 5.4c). Absorption at 498, 390 and 280 nm were monitored in order to determine the proportion of well-folded pSRII-KCWL with respect to the total protein concentration of each eluted fraction. Each eluted fraction was analysed by UV/vis spectroscopy and SDS-PAGE (Figure 5.4d–e). Fractions 9–11 showed predominantly well-folded pSRII-KCWL, with minimal amounts of misfolded protein. The proportion of misfolded protein was higher in later fractions, as evidenced by the increasing intensity of the 390-nm peak. The two species could not be completely separated. The later elution of misfolded species suggests that, on average, fewer DDM monomers surround these species compared to well-folded protein. This could perhaps be attributed to the low DDM concentration (0.05%) being used in the gel filtration buffer.

To determine the stability of pSRII-KCWL, the protein was incubated at 25, 35 and 45 °C for 16 h, showing nearly identical UV/vis spectra before and after incubation. This indicates that pSRII-KCWL was stable under the examined conditions, thus providing a benchmark for assessing the stability of *trans*-spliced pSRII-KCWL.

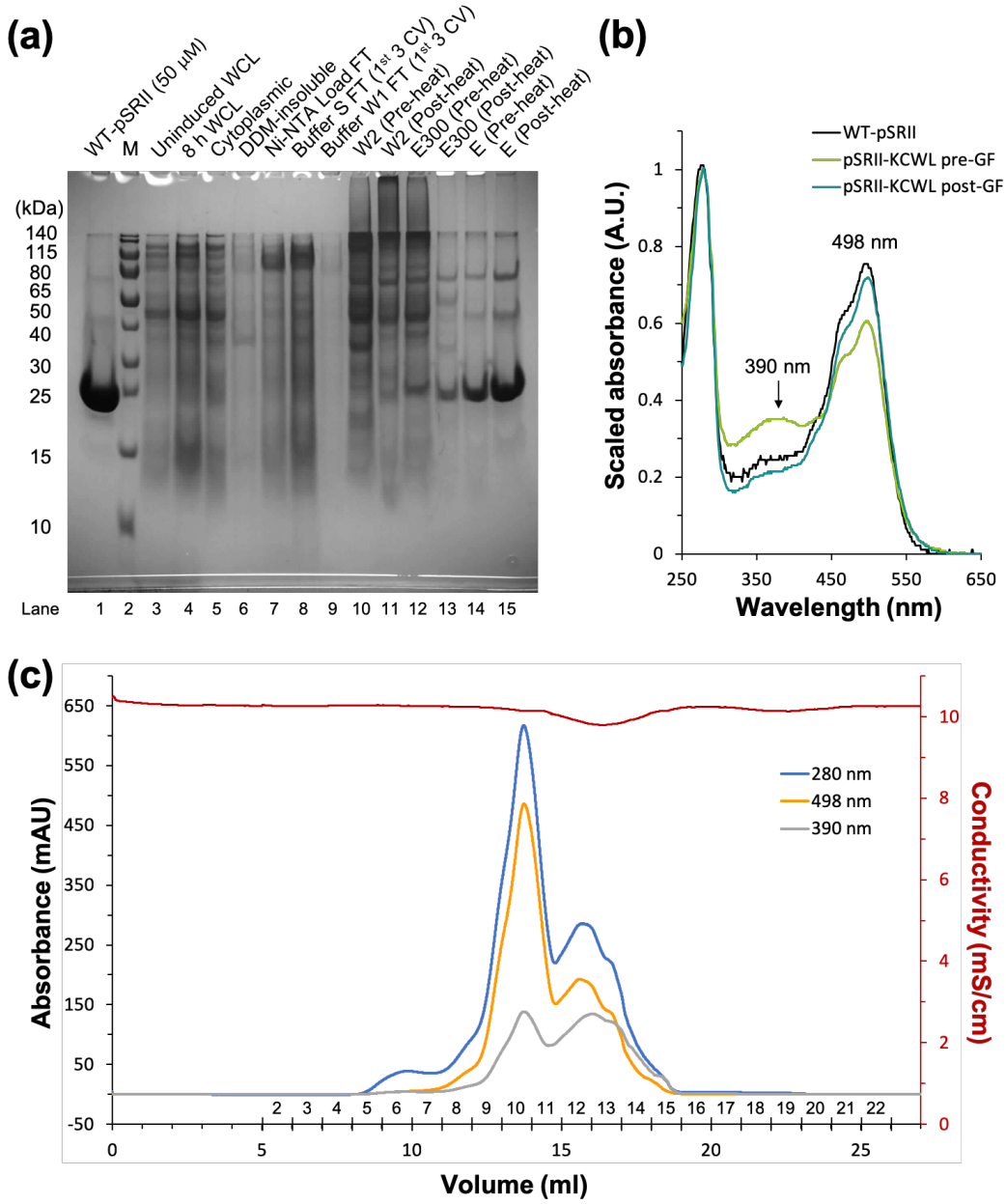


Figure 5.4: Purification of full-length pSRII-KCWL
(a) SDS-PAGE showing the expression and purification of full-length pSRII-KCWL. Lane 1, 50 µM WT-pSRII in c7-DHPC; lane 2, molecular weight marker (M); lane 3, whole cell lysate (WCL) of uninduced Tuner(DE3)LacI cells; lane 4, WCL after pSRII-KCWL was expressed for 8 h at 30 °C; lane 5, cytoplasmic fraction of lysed culture; lane 6, DDM-insoluble fraction solubilised using 3% SDS + 9 M urea; lane 7, fraction which could not be bound to Ni-NTA beads; lanes 8–9, flow-through from washing pSRII-KCWL bound to Ni-NTA beads with Buffer S (lane 8) and Buffer W1 (lane 9); elution with Buffer W2 (lanes 10–11), Buffer E300 (lanes 12–13) and Buffer E (lanes 14–15) before and after heat incubation. **(b)** UV/vis spectra of WT-pSRII (black) and pSRII-KCWL before (lime) and after (teal) size exclusion chromatography, scaled according to absorbance at 280 nm. **(c)** Size exclusion chromatography trace for pSRII-KCWL purification using a Superdex 200 Increase 10/300 GL column running at 0.5 ml/min at 4 °C. Each eluted fraction was 1 ml.

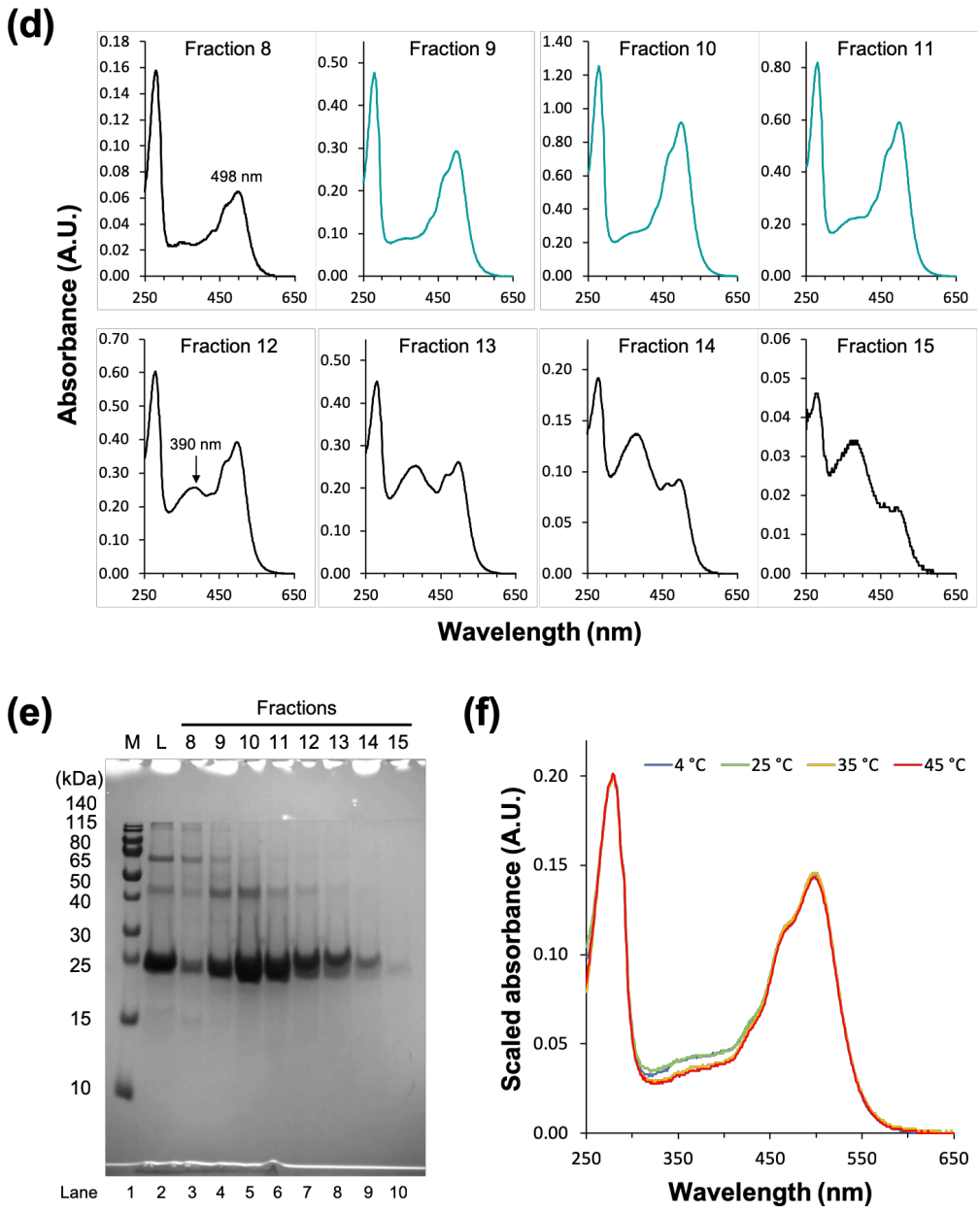


Figure 5.4: Purification of full-length pSRII-KCWL (cont.)
(d) UV/vis spectra of fractions 8–15 eluted from size exclusion chromatography of pSRII-KCWL. Fractions 9–11 had the greatest proportions of well-folded pSRII-KCWL (λ_{max} = 498 nm) and little misfolded species (λ_{max} = 390 nm). The proportion of misfolded species increased in later fractions. **(e)** SDS-PAGE gel for fractions 8−15. M: molecular weight marker; L: pre-gel filtration sample. **(f)** UV/vis spectra of pSRII-KCWL in DDM incubated at different temperatures for 16 h, demonstrating the stability of pSRII-KCWL. The spectra were scaled according to absorbance at 280 nm.

5.2.3 Test expression

To maximise the yield of the ligated product, individual test expression of pSRII^{N}(KCWL)-DnaE^{N}-Strep and Strep-DnaE^{C}-pSRII^{C}(KCWL)-His$_6$ were conducted to identify the optimal temperature and inducer concentration for expressing the two precursors.

For Strep-DnaE^{C}-pSRII^{C}(KCWL)-His$_6$, expression at 18, 25 and 37 °C, both in the absence and presence of retinal, were first attempted (Figure 5.5). Whole cell lysates were analysed by SDS-PAGE and Western blotting using HisProbe-HRP conjugate. The whole cell lysate of an uninduced sample showed no expression. Fast onset of expression was observed within 1 h after addition of 0.1% L-arabinose. Test expression was monitored for up to 20 h to determine the stability of the C-precursor, should prolonged expression times be required for maximal yield of ligated product in dual-expression of both precursors. Overnight expression at 37 °C led to significant degradation of the expressed C-precursor, while similar yields were obtained at 18 and 25 °C. Most of the C-precursor localised to the crude membrane fraction (Figure 5.6). Some of the protein expressed at 18 °C could be observed in the soluble fraction, although it is unclear whether this is due to *bona fide* partial localisation to the cytosol or due to the protein being solubilised by detergents in the RIPA buffer. Similar results were obtained regardless of the absence or presence of all-*trans* retinal during expression. The expression levels of the C-precursor upon induction by adding 0.05, 0.1 and 0.2% L-arabinose at 18 °C were also assessed (Figure 5.7). A gradual accumulation of the C-precursor was observed for all conditions, with induction by 0.05% L-arabinose yielding the fastest accumulation of the C-precursor. Although an N-terminal Strep-tag II was cloned into the construct, the StrepTactin-HRP probe did not yield very visible bands despite the probe being reported as suitable for detecting N-terminal, C-terminal and internal Strep-tag II. This suggests that either the expression levels of the C-precursor were not ideal despite extensive attempts at optimisation, or the binding of StrepTactin-HRP probe to the N-terminal Strep-tag II was weak. It would therefore be impractical to attempt detecting protein *trans*-splicing by monitoring the depletion of the C-precursor.

Expression of pSRII^{N}(KCWL)-DnaE^{N}-Strep was first monitored at 18, 25 and 30 °C for up to 20 h (Figure 5.8a–b). Higher temperatures were not tested given the low yield of the C-precursor after overnight expression at 37 °C. All-*trans* retinal was not added in expression

trials of the N-precursor, since the absence of Lys-205 would preclude formation of the retinal Schiff base. Western blot analysis shows very small amount of leaky expression in the uninduced whole cell lysate, and fast onset of N-precursor expression within 1 h after adding 1 mM IPTG. Out of the three temperatures, overnight expression at 18 °C gave the greatest yield. The N-precursor appeared to oligomerise, regardless of lysing the *E. coli* culture by sonication in Buffer A with 5 mM TCEP-HCl (Figure 5.8a) or using RIPA buffer (Figure 5.8b) (see Materials & Methods). This indicates that oligomerisation was not due to the presence of SDS in RIPA buffer (cf. oligomerisation of SDS-denatured pSRII in Chapter 2) or formation of disulphide bonds between protomers. Despite the presence of oligomeric species, there was a higher proportion of monomeric N-precursor which should remain accessible for *trans*-splicing. The effects of using different concentrations of IPTG (0.1, 0.3 and 1 mM) for inducing N-precursor expression at 18 °C were also studied. There was little difference in expression yield across the three conditions, although 0.1 mM IPTG gave arguably slightly greater yield. Almost all of the expressed N-precursor localised to the crude membrane fraction (Figure 5.8c).

In summary, out of the conditions examined, expression of Strep-DnaE^{C}-pSRII^{C}(KCWL)-His_6 was optimal at 18 °C in the presence of 0.05% L-arabinose, and expression of pSRII^{N}(KCWL)-DnaE^{N}-Strep was optimal at 18 °C in the presence of 0.1 mM IPTG. Both precursors localised to the membrane fraction.

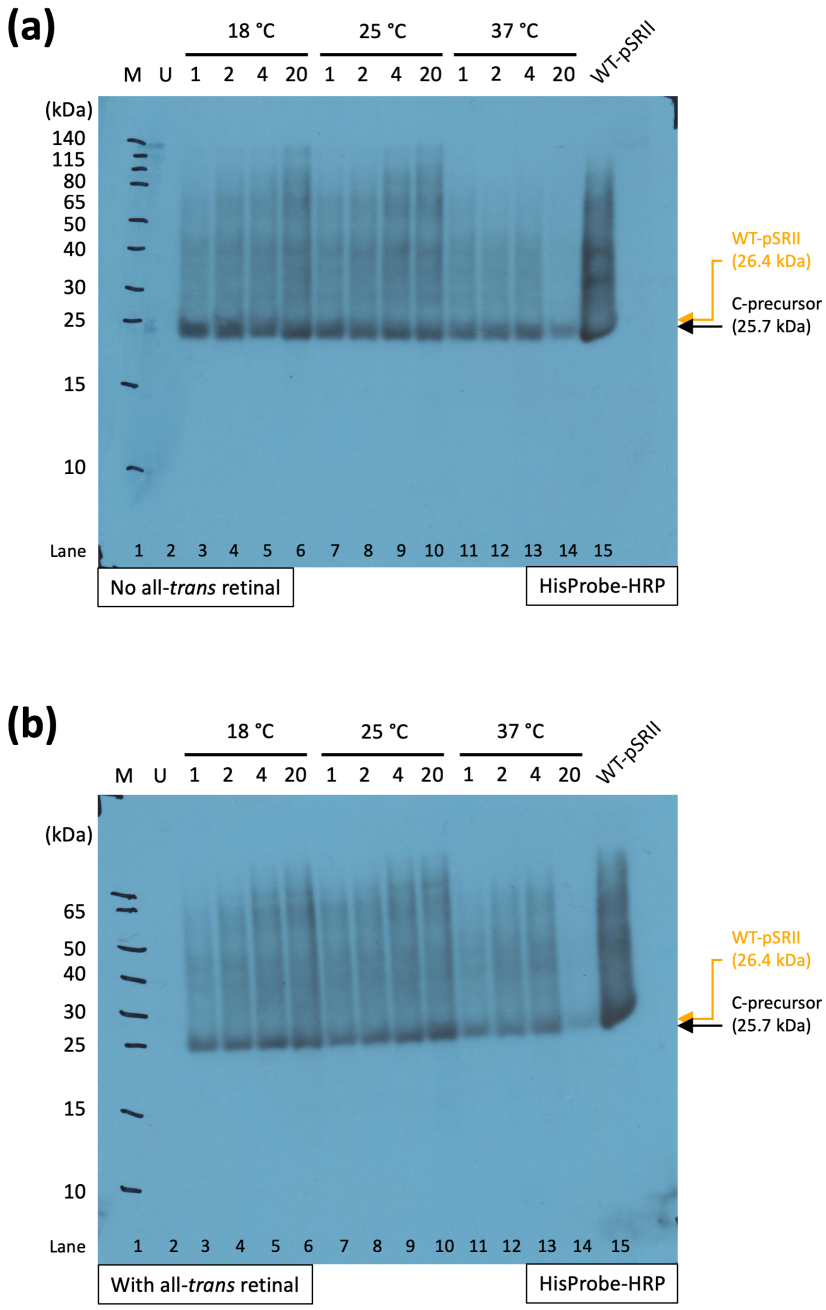


Figure 5.5: Optimising expression temperature for Strep-DnaE^{C}-pSRII^{C}(KCWL)-His$_6$
The C-precursor was expressed in Tuner(DE3)LacI cells at different temperatures (18, 25 and 37 °C), both with **(a)** and without **(b)** supplementation of all-*trans* retinal at 2-h intervals. Expression was induced by the addition of 0.1% L-arabinose at OD_{600} = 0.8–1.0, and monitored for up to 20 h. The C-precursor was detected from whole cell lysates by Western blotting using HisProbe-HRP. M: molecular weight marker; U: uninduced whole cell lysate; positive control: 0.1 µM WT-pSRII in c7-DHPC. The arrows indicate the expected migration of the C-precursor and WT-pSRII.

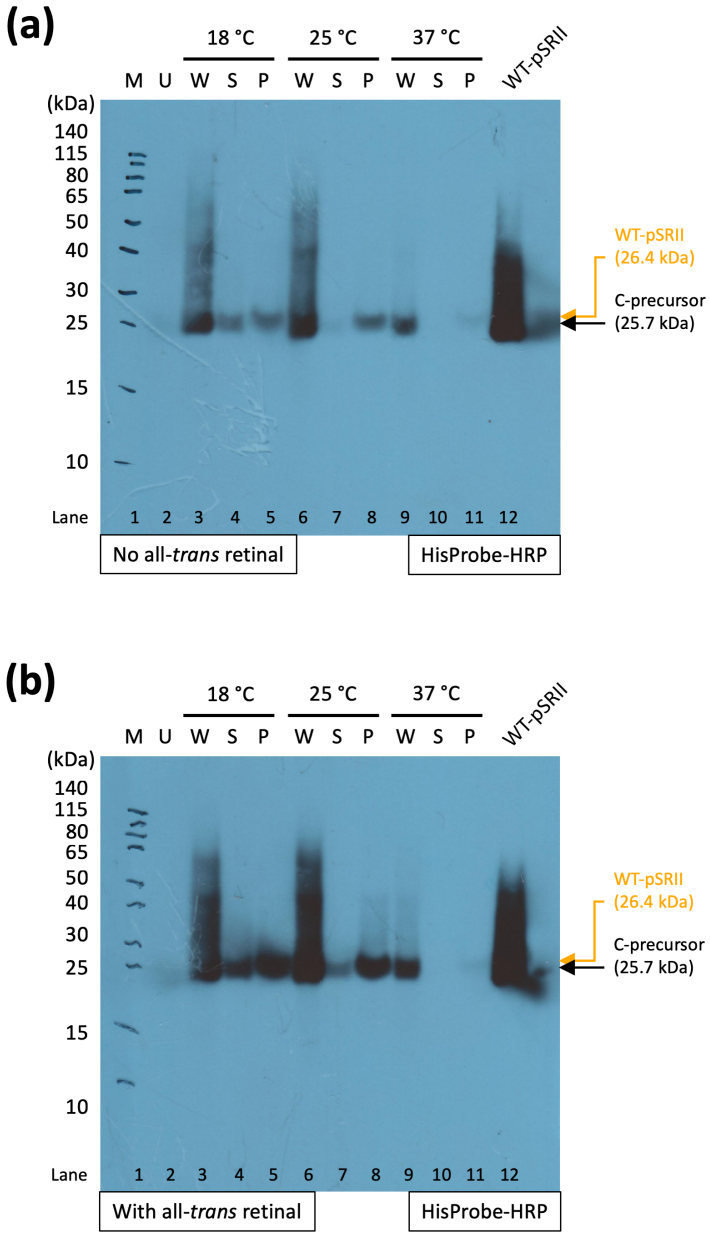


Figure 5.6: Localisation of Strep-DnaE^{C}-pSRII^{C}(KCWL)-His$_6$
The C-precursor was expressed in Tuner(DE3)LacI cells at different temperatures (18, 25 and 37 °C), both with **(a)** and without **(b)** supplementation of all-*trans* retinal at 2-h intervals. The C-precursor was detected from whole cell lysates (W), soluble fraction (S) and crude membranes (P) by Western blotting using HisProbe-HRP. M: molecular weight marker; U: uninduced whole cell lysate; positive control: 0.1 µM WT-pSRII in c7-DHPC. The arrows indicate the expected migration of the C-precursor and WT-pSRII.

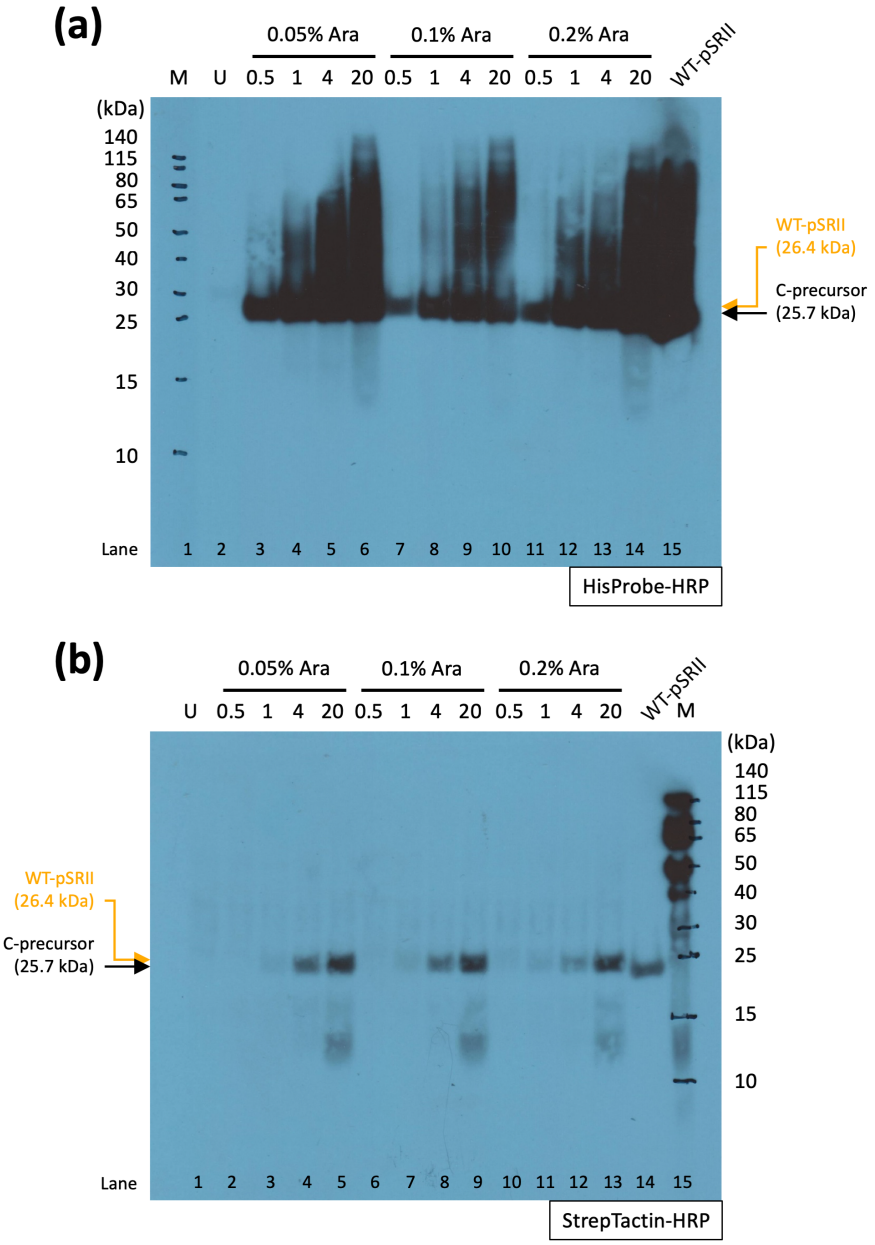


Figure 5.7: Optimising inducer concentration for Strep-DnaE^{C}-pSRII^{C}(KCWL)-His$_6$
The C-precursor was expressed in Tuner(DE3)LacI cells at 18 °C, with supplementation of all-*trans* retinal at 2-h intervals. Expression was induced by adding 0.05, 0.1 or 0.2% L-arabinose, and monitored for up to 20 h. The C-precursor was detected from whole cell lysates by Western blotting using HisProbe-HRP **(a)** and StrepTactin-HRP **(b)**. M: molecular weight marker; U: uninduced whole cell lysate. 0.1 µM WT-pSRII in c7-DHPC was used as the positive control in panel a, and as the negative control in panel b. The arrows indicate the expected migration of the C-precursor and WT-pSRII.

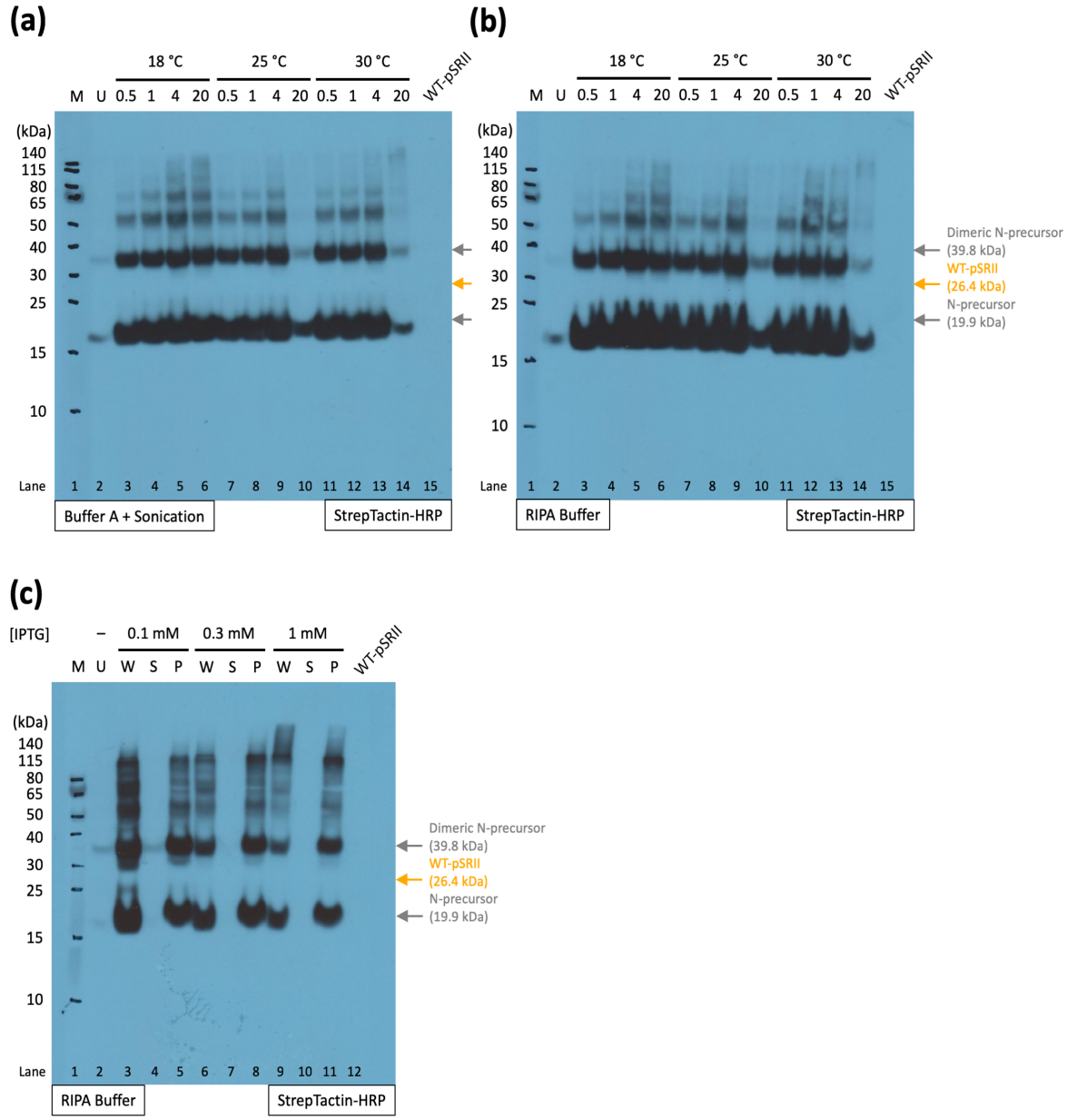


Figure 5.8: Optimising expression conditions of pSRII^{N}(KCWL)-DnaE^{N}-Strep
(a–b) Expression of the N-precursor in Tuner(DE3)LacI cells was induced by adding 1 mM IPTG. Expression at different temperatures (18, 25 and 30 °C) was monitored for up to 20 h. Cells were lysed by sonication in Buffer A **(a)** or by incubation in RIPA buffer **(b)**, and expression levels were assessed using whole cell lysates. **(c)** Expression at 18 °C was induced by adding different concentrations of IPTG (0.1, 0.3 and 1 mM). Expression yield and localisation of the N-precursor were assessed using whole cell lysates (W), soluble fraction (S) and crude membranes (P). For all panels, the N-precursor was detected by Western blotting using StrepTactin-HRP. M: molecular weight marker; U: uninduced whole cell lysate; negative control: 0.1 µM WT-pSRII in c7-DHPC. The arrows indicate the expected migration of the N-precursor, WT-pSRII and dimeric N-precursor.

5.2.4 Dual-expression

Dual-expression of the N- and C-precursors was attempted using the optimal expression conditions identified from individual test expressions of the two precursors. Four conditions were tested to further optimise dual-expression (Table 5.1). Several factors which might affect the yield of ligated pSRII-KCWL were examined, including the presence (condition 1) vs. absence (condition 2) of all-*trans* retinal, the order in which expression of the two precursors were induced (condition 3), and expression temperature (condition 4). Given the stable membrane localisation of both precursors, the order of expression could be important for determining the final yield of the ligated product. Ideally, the precursor which localises slower to the membrane should be expressed first, such that this precursor is available within the cytosol for *trans*-splicing once the other precursor is expressed.

Condition	Precursor being expressed first	Temperature (°C)	All-*trans* retinal
1	C-precursor	18	✓
2	C-precursor	18	-
3	N-precursor	18	✓ only upon expression of C-precursor
4	C-precursor	25 °C for 8 h 18 °C overnight	✓

Table 5.1: Conditions tested for dual-expression of pSRII^{N}(KCWL)-DnaE^{N}-Strep and Strep-DnaE^{C}-pSRII^{C}(KCWL)-His$_6$

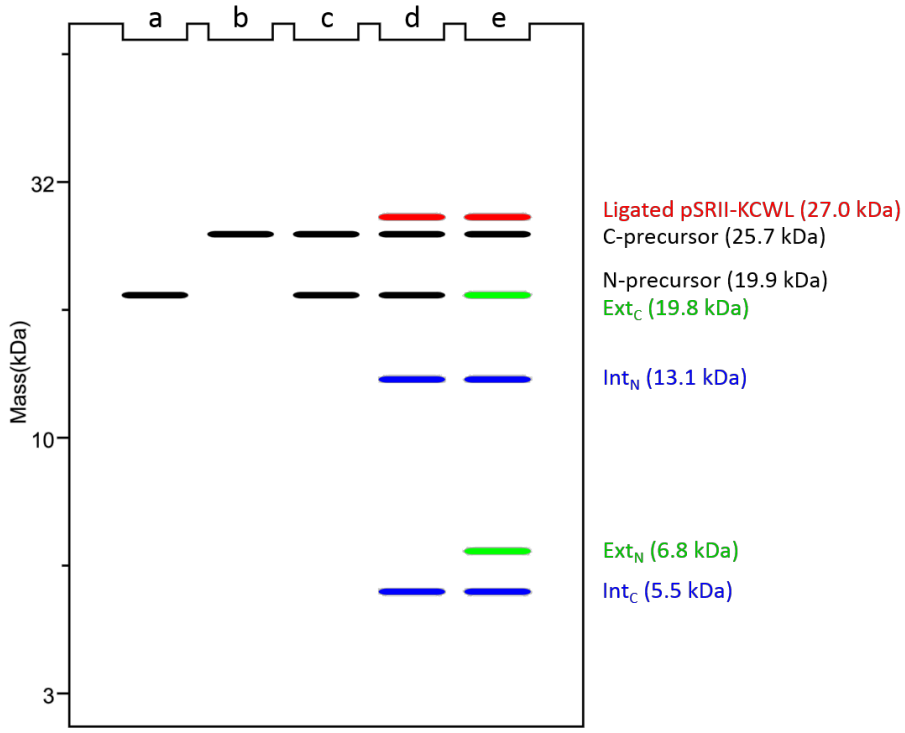


Figure 5.9: Schematic illustration of different patterns in SDS-PAGE upon dual-expression. **(a)** N-precursor only. **(b)** C-precursor only. **(c)** Dual-expression without successful *trans*-splicing. **(d)** Dual-expression with successful *trans*-splicing. **(e)** Dual-expression with *trans*-splicing and *trans*-cleavage. Figure was produced using ChemDraw (PerkinElmer).

Aliquots were taken from cell cultures at specific time-points, and expression yield was analysed by probing for His_6-tagged and Strep-tagged proteins on Western blots. In principle, successful protein *trans*-splicing can be assessed from the appearance of additional bands which are different from the two precursor proteins (Figure 5.9d). If splicing occurs with 100% efficiency and equal amounts of both precursors, only one new dominant band corresponding to the ligated product would be observed and all precursors would be depleted. In practice, it is not trivial to obtain equal expression of both precursors, so at least one of the precursors would still be observable on SDS-PAGE. Detection of ligated pSRII-KCWL was difficult. The C-precursor (25.7 kDa) and ligated pSRII-KCWL (27.0 kDa) have very similar molecular weights (Figure 5.9d), thus hampering straightforward detection of the ligated product by probing for His_6-tagged protein on Western blot. The Strep-tagged C-intein (excised upon successful *trans*-splicing) has molecular weight of 5.5 kDa, which would have been too small to be detected by 12% SDS-PAGE. In principle, the difference in molecular weight between the excised Strep-tagged N-intein (13.1 kDa) and the N-precursor (19.9 kDa)

could provide sufficient resolution of the two species by SDS-PAGE. However, no bands were observed below 15 kDa, possibly indicating that *trans*-splicing was not successful. On the other hand, the high expression levels of the N-precursor yielded very thick bands on the Western blots, and it is possible for membrane proteins to migrate anomalously on SDS-PAGE compared to their expected molecular weights.

Greater amounts of leaky expression were observed in the uninduced whole cell lysates across all four different expression conditions. There was little difference in expression yield in the presence (condition 1) vs. absence (condition 2) of all-*trans* retinal (Figure 5.10a–b). This is somewhat surprising, considering the important roles of all-*trans* retinal in backbone amide dynamics and the tertiary structure of pSRII (see Chapters 2–3). Similar expression yields were obtained across the three conditions in which expression of the C-precursor was induced first, while induction of N-precursor expression prior to C-precursor expression (condition 3) gave slightly lower yield (Figure 5.10c–d).

To summarise, dual-expression of Strep-DnaE^{C}-pSRII^{C}(KCWL)-His$_6$ and pSRII^{N}(KCWL)-DnaE^{N}-Strep showed expression of each precursor. The inability to observe any excised N-intein possibly indicates that *trans*-splicing might not have occurred, although this conclusion could not be made with great certainty given the similar molecular weights of the precursors, ligated product and side-products. Expression yield appeared to be slightly higher if the C-precursor was expressed prior to the N-precursor, although all four conditions examined resulted in fairly comparable expression yields.

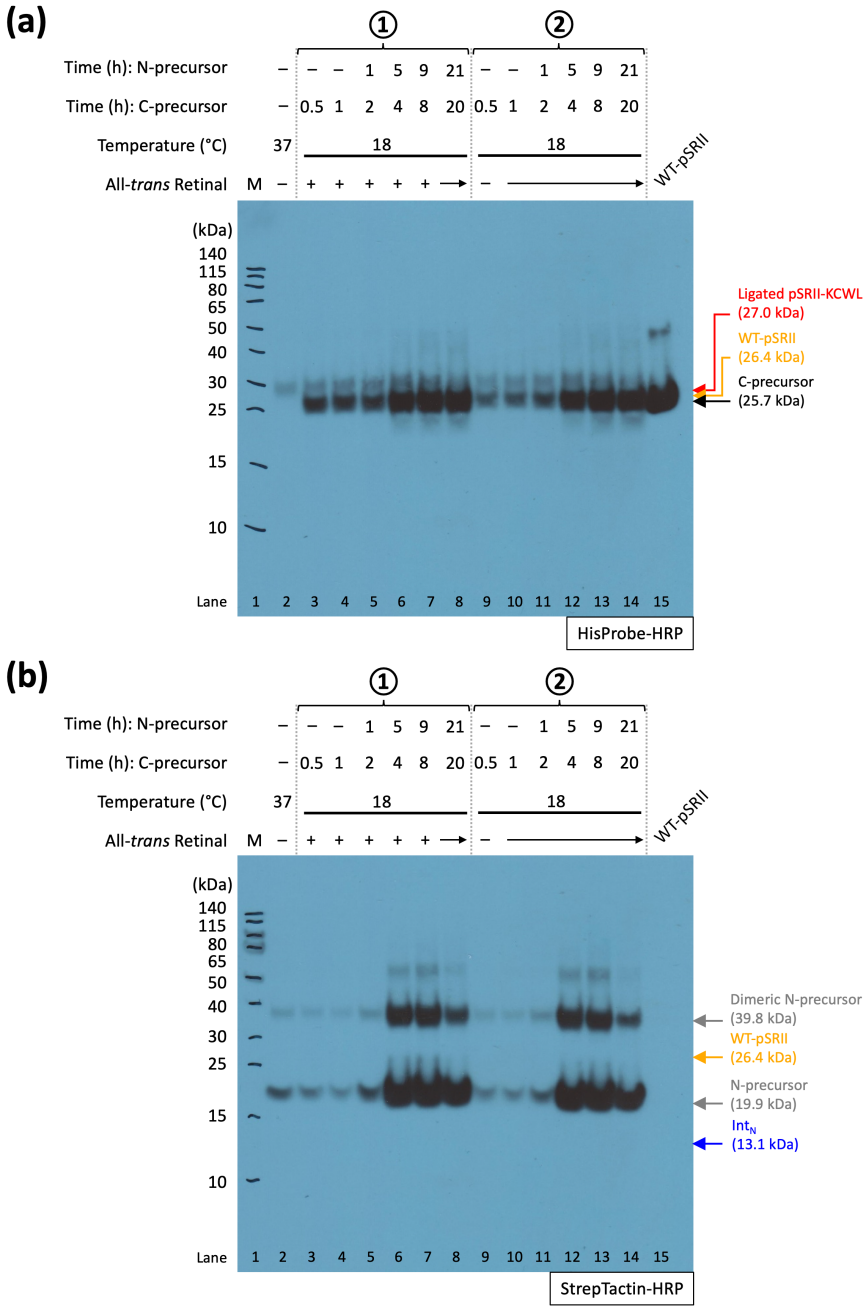


Figure 5.10: Dual-expression of pSRII^N(KCWL)-DnaE^N-Strep and Strep-DnaE^C-pSRII^C(KCWL)-His$_6$
(a–b) Expression of the C-precursor in Tuner(DE3)LacI cells was induced by adding 0.05% L-arabinose before expression of the N-precursor was induced by adding 0.1 mM IPTG. Expression was conducted at 18 °C in the presence (Condition 1) vs. absence (Condition 2) of all-*trans* retinal, and was monitored for up to 21 h. M: molecular weight marker; U: uninduced whole cell lysate. 0.1 µM WT-pSRII in c7-DHPC was used as the positive control in panel a and as the negative control in panel b. In panel a, the arrows indicate the expected migration of the C-precursor, WT-pSRII and ligated pSRII-KCWL. In panel b, the arrows indicate the expected migration of the cleaved N-intein (Int$_N$), the N-precursor, WT-pSRII and dimeric N-precursor.

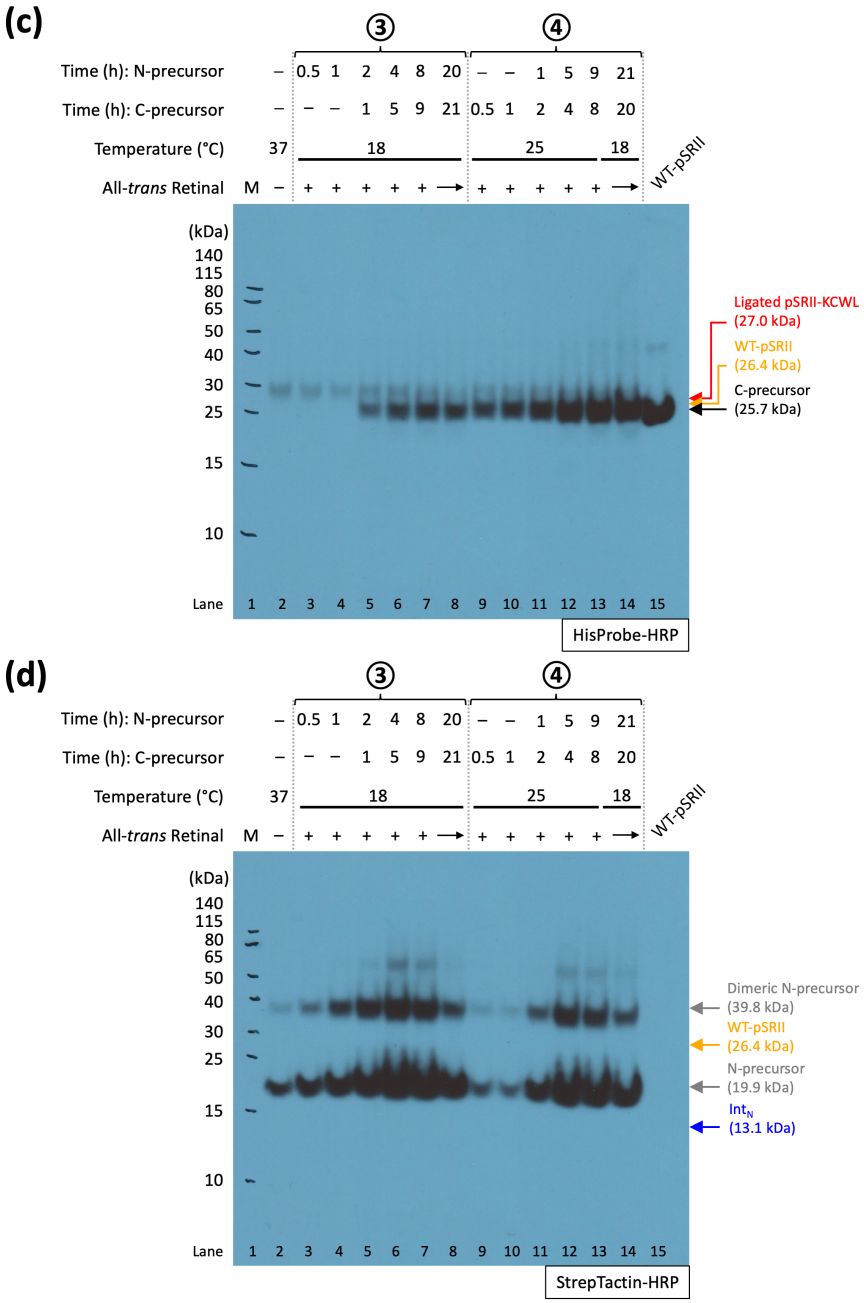


Figure 5.10: Dual-expression of pSRII^N(KCWL)-DnaE^N-Strep and Strep-DnaE^C-pSRII^C(KCWL)-His$_6$ (cont.)

(c–d) Condition 3: Expression of the N-precursor in Tuner(DE3)LacI cells was induced by adding 0.1 mM IPTG before expression of the C-precursor was induced by adding 0.05% L-arabinose. Expression was conducted at 18 °C in the presence of all-*trans* retinal for 21 h. Condition 4: Expression of the C-precursor in Tuner(DE3)LacI cells was induced by adding 0.05% L-arabinose before expression of the N-precursor was induced by adding 0.1 mM IPTG. Expression was conducted at 25 °C for the first 9 h and at 18 °C for the next 12 h. M: molecular weight marker; U: uninduced whole cell lysate. 0.1 µM WT-pSRII in c7-DHPC was used as the positive control in panel c and as the negative control in panel d. In panel c, the arrows indicate the expected migration of the C-precursor, WT-pSRII and ligated pSRII-KCWL. In panel d, the arrows indicate the expected migration of the cleaved N-intein (Int$_N$), the N-precursor, WT-pSRII and dimeric N-precursor.

Considering the aforementioned complications with detecting the ligated pSRII-KCWL, dual-expression of pSRII^{N}(KCWL)-DnaE^{N}-Strep and Strep-DnaE^{C}-pSRII^{C}(KCWL)-His$_6$ was performed with 3 L of *E. coli* Tuner(DE3)LacI culture (Figure 5.11). Expression was conducted at 25 °C for the first 8 h and at 18 °C overnight. Expression of the C-precursor was induced first with 0.05% L-arabinose followed by induction of N-precursor expression with 0.1 mM IPTG. All-*trans* retinal was supplemented at 2-h intervals for the first 8 h.

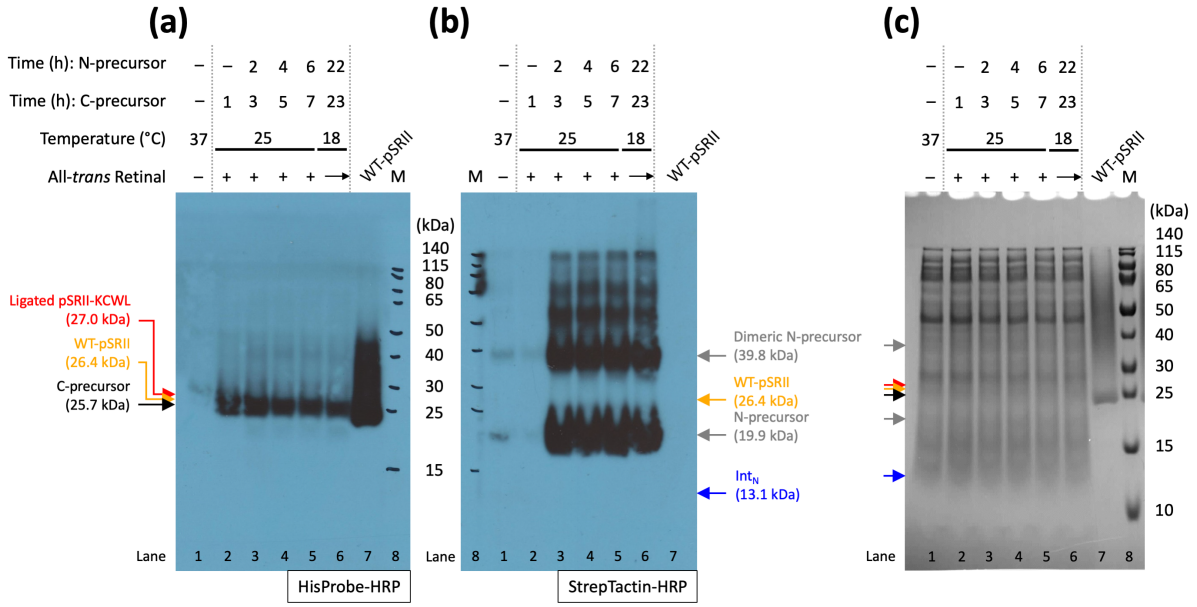


Figure 5.11: Large-scale dual-expression in *E. coli* Tuner(DE3)LacI cells
Whole cell lysates from specific time-points of overnight dual-expression were analysed by Western blotting using HisProbe-HRP **(a)** and StrepTactin-HRP **(b)**, and by SDS-PAGE **(c)**. M: molecular weight marker. 0.1 µM WT-pSRII in c7-DHPC was used as the positive control in panel a and as the negative control in panel b. In panel a, the arrows indicate the expected migration of the C-precursor, WT-pSRII and ligated pSRII-KCWL. In panel b, the arrows indicate the expected migration of the cleaved N-intein (Int$_N$), the N-precursor, WT-pSRII and dimeric N-precursor. In panel c, the arrows indicate the expected migration of all six species which are marked in panels a and b (from low to high molecular weight: Int$_N$, N-precursor, C-precursor, WT-pSRII, ligated pSRII-KCWL and dimeric N-precursor).

Purification of putative ligated pSRII-KCWL was performed using the same method as for the full-length pSRII-KCWL mutant (see section 5.2.2), and aliquots taken from each stage of the purification protocol were analysed by SDS-PAGE and Western blotting (Figure 5.12a–c). The yield was alarmingly low; bands corresponding to the expected species were unobservable on SDS-PAGE even after concentrating the eluted fractions (W2, E and

E300) using a 3 kDa MWCO concentrator, and must be detected by Western blotting. One dominant band running at ~25 kDa was observed in fractions W2, E and E300 on the Western blot detected with HisProbe-HRP, likely to be corresponding to the C-precursor. A faint band could be observed just below the dominant band, which could suggest a small amount of C-terminal cleavage leading to liberation of the C-extein (Figure 5.12a).

A UV/vis absorption spectrum was recorded for fraction E, in which most of the protein was eluted. The spectrum showed clear absence of the 498-nm peak corresponding to all-*trans* retinal in a properly-folded and ligated pSRII-KCWL, providing conclusive evidence that protein *trans*-splicing was not successful in generating functional protein. The peaks detected in the 300- to 650-nm spectral region likely corresponded to different isomers or chemical modifications of retinal. It is noted that the 559-nm peak is significantly red-shifted compared to the reported absorption wavelengths for isomers and chemical derivatives of retinal in ethanol [240], and is instead closer to the range of different wavelengths reported for visual pigments [418]. Although it might be possible for protein *trans*-splicing and retinal Schiff base formation to have occurred without formation of a functional retinal binding pocket, this would be very difficult to assess by biophysical techniques or mass spectrometry due to the extremely low yield.

The split intein method can, in principle, be applied both *in vitro* and *in vivo*. However, Western blot analysis shows that the C-precursor is DDM-soluble but the N-precursor is not (Figure 5.12b). This indicates that for future studies involving *in vitro* assembly of pSRII, different detergents and membrane mimetics must be screened to identify conditions which are suitable for (1) solubilising both precursors, (2) maintaining both precursors in conformations that are amenable for *trans*-splicing, and (3) maintaining the functionality of ligated pSRII-KCWL.

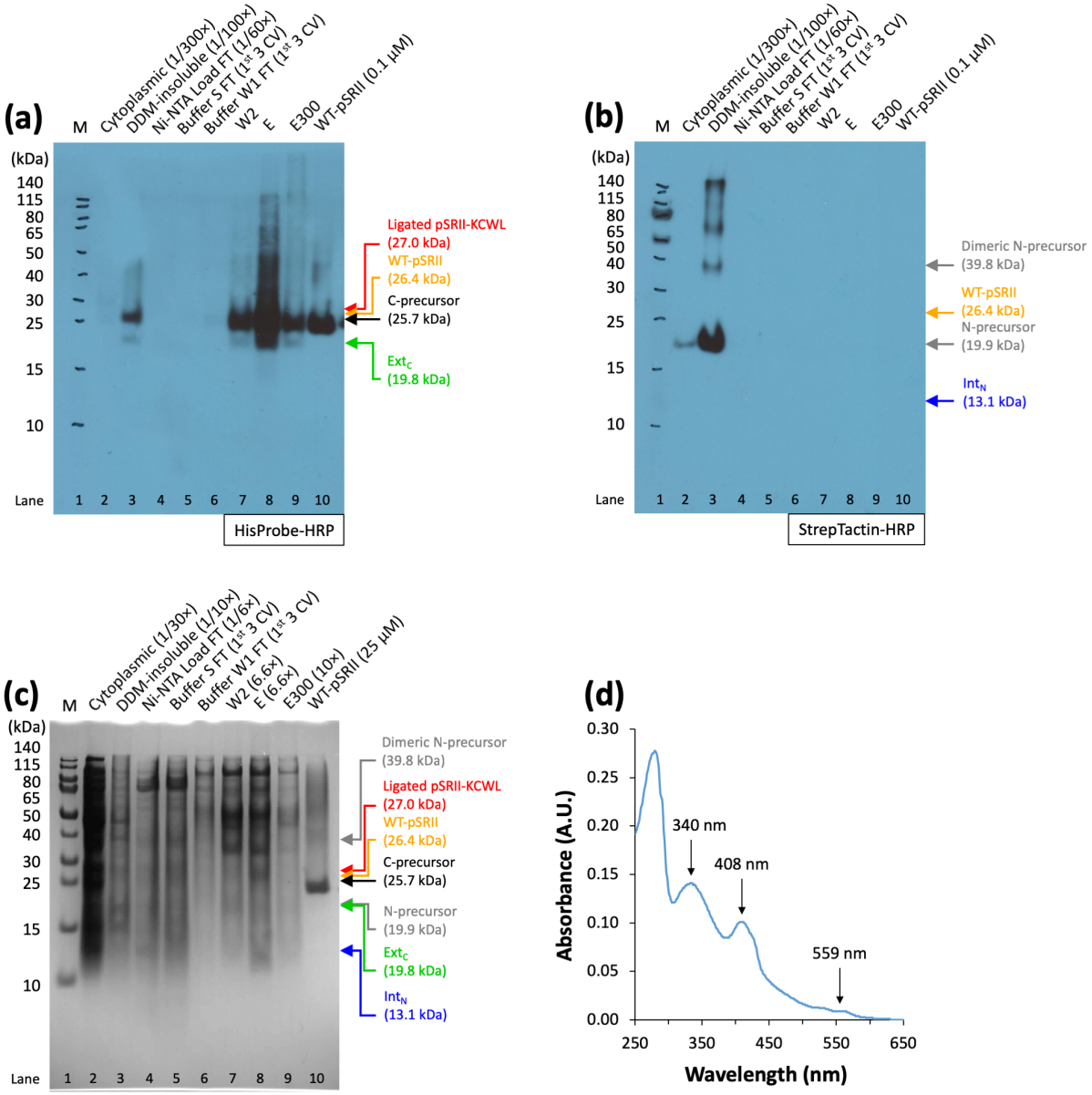


Figure 5.12: Purification from *E. coli* Tuner(DE3)LacI cells
Aliquots were taken from each stage of purification and analysed by Western blotting using HisProbe-HRP **(a)** and StrepTactin-HRP **(b)**, and by SDS-PAGE **(c)**. Lane 1, molecular weight marker (M); lane 2, cytoplasmic fraction of lysed culture; lane 3, DDM-insoluble fraction solubilised using 3% SDS + 9 M urea; lane 4, fraction which could not be bound to Ni-NTA beads; lanes 5–6, flow-through from washing pSRII-KCWL bound to Ni-NTA beads with Buffer S (lane 5) and Buffer W1 (lane 6); lanes 7–9, elution with Buffer W2, Buffer E and Buffer E300, respectively; lane 10, WT-pSRII in c7-DHPC. In panel a, the arrows indicate the expected migration of the cleaved C-extein (Ext$_C$), the C-precursor, WT-pSRII and ligated pSRII-KCWL. In panel b, the arrows indicate the expected migration of the cleaved N-intein (Int$_N$), the N-precursor, WT-pSRII and dimeric N-precursor. In panel c, the arrows indicate the expected migration of all seven species which are marked in panels a and b (from low to high molecular weight: Int$_N$, Ext$_C$, N-precursor, C-precursor, WT-pSRII, ligated pSRII-KCWL and dimeric N-precursor). **(d)** UV/vis spectrum of Fraction E, showing clear absence of the 498-nm peak expected for well-folded pSRII-KCWL.

The C41(DE3) *E. coli* strain is typically used for producing proteins which express poorly or are toxic for BL21(DE3). C41(DE3) is suitable for the overexpression of many membrane proteins, although the reason for its improved characteristics is not well-understood [419,420].

Dual-expression of the N- and C-precursors in C41(DE3) cells and purification were performed as described previously. The onset of N-precursor expression was noted to be slower in C41(DE3) compared to that in Tuner(DE3)LacI cells, as the N-precursor was barely detectable by Western blotting even at 2 h after induction (Figure 5.13b). The N-precursor was predominantly DDM-insoluble, but a small fraction appeared to have co-purified with the C-precursor. There was extra complication with degradation of the C-precursor beginning from between 3 and 7 h after its expression was induced. This suggests that the limiting factor of this experiment could have been the low amounts of the C-precursor available after overnight expression.

Similar to dual-expression in Tuner(DE3)LacI cells, the final yield in fraction E was also very low, as the expected bands were difficult to detect on SDS-PAGE. The 498-nm peak indicative of well-folded ligated pSRII-KCWL could not be observed either (Figure 5.13d). The species eluted in W2, E and E300 fractions would therefore be predominantly the C-precursor, with a faint band at ~20 kDa which likely corresponds to the cleaved C-extein (Figure 5.13a). Interestingly, an extra band running between 30 and 40 kDa was observed in Fraction E detected by HisProbe-HRP. The molecular weight of this band does not match any of the expected products (Figure 5.9). There are two possible candidates for this species: the branched intermediate (expected molecular weight = 32.5 kDa) or dimeric C-extein (expected molecular weight = 39.6 kDa). Observation of the branched intermediate is more likely, and would suggest that protein *trans*-splicing was limited by excision of the split intein, which should have occurred by cyclisation of Asn, the last residue of the C-intein. The stability of the pSRII branched intermediate in the cellular context and during purification are unknown, but its observation has been reported for *trans*-splicing mediated by the Psp pol intein-1 [391]. The less plausible explanation for the extra band is the observation of dimeric C-extein, which could suggest that significantly more C-terminal cleavage has occurred during expression in C41(DE3) cells, and the liberated C-extein has a tendency to dimerise.

However, the reason behind different amounts of C-terminal cleavage in Tuner(DE3)LacI and C41(DE3) cells is unknown.

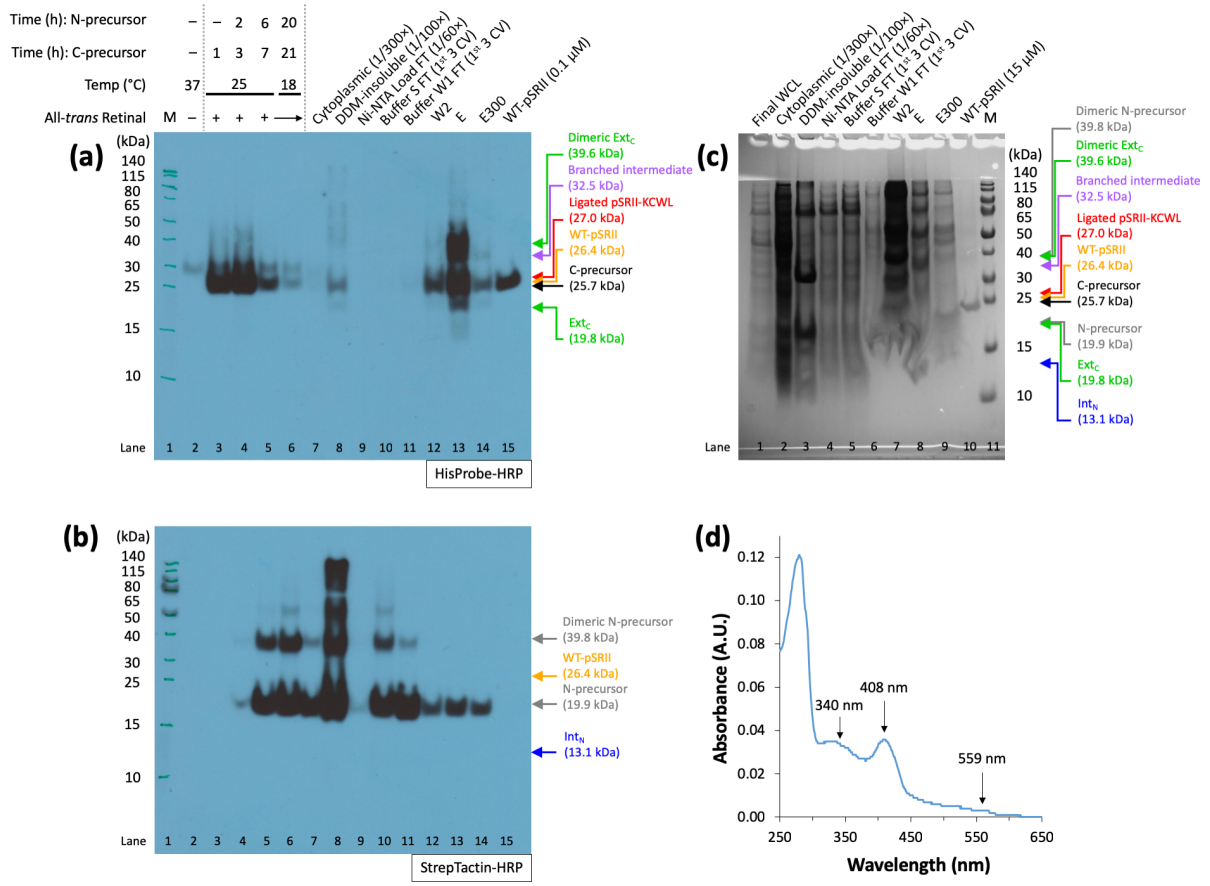


Figure 5.13: Large-scale dual-expression in and purification from in *E. coli* C41(DE3) cells
Aliquots were taken from specific time-points during expression and at each stage of purification, and analysed by Western blotting using HisProbe-HRP **(a)** and StrepTactin-HRP **(b)**, and by SDS-PAGE **(c)**. In panel a, the arrows indicate the expected migration of the cleaved C-extein (Ext$_C$), the C-precursor, WT-pSRII, ligated pSRII-KCWL, branched intermediate and dimeric Ext$_C$. In panel b, the arrows indicate the expected migration of the cleaved N-intein (Int$_N$), the N-precursor, WT-pSRII and dimeric N-precursor. In panel c, the arrows indicate the expected migration of all nine species which are marked in panels a and b (from low to high molecular weight: Int$_N$, Ext$_C$, N-precursor, C-precursor, WT-pSRII, ligated pSRII-KCWL, branched intermediate, dimeric Ext$_C$ and dimeric N-precursor). **(d)** UV/vis spectrum of fraction E, showing clear absence of the 498-nm peak expected for well-folded pSRII-KCWL.

5.3 Discussion and conclusions

In this chapter, attempts were made to assemble pSRII using the *Npu* DnaE split intein, with the vision of applying this method towards studying the roles of different loops and helices in the folding pathway of pSRII. Protein *trans*-splicing *in vivo* was chosen because it permits protein assembly from complementary fragments without the need for optimising reaction conditions for the splicing reaction or separate purification protocols for each precursor, as would have been required for *in vitro* approaches. The splice site was introduced between the two β-strands in the BC loop of pSRII. The relevant constructs were cloned to encode for helices A–B fused with the N-terminal half of the DnaE split intein (the N-precursor) and for helices C–G fused with the C-terminal half of the split intein (the C-precursor). Numerous attempts were made to optimise the conditions for individual expression of each precursor and for dual-expression of both precursors using *E. coli* Tuner(DE3)LacI cells. However, large-scale dual-expression in Tuner(DE3)LacI and C41(DE3) cells did not yield well-folded pSRII-KCWL, as evidenced by the lack of absorption peak at 498 nm.

The most important factors contributing towards successful utilisation of split intein tools are: (1) the yield of the ligated product, (2) the rate constant of the *trans*-splicing reaction, (3) the solubility of the precursor fragments, and (4) tolerance for non-native extein sequences [401]. The observations reported in this chapter can be reconciled with potential problems in the above factors occurring at various steps of the *trans*-splicing mechanism. Firstly, association of the two precursors was likely to have been a major limiting factor, as evidenced by the observation of both the N- and C-precursors but not the excised N-intein indicative of successful *trans*-splicing during small-scale and large-scale dual-expression. The fast rate of the *trans*-splicing reaction [396] suggests that *trans*-splicing of pSRII was unlikely to be limited by reaction kinetics. Instead, it is hypothesised here that very minimal amounts of the two precursors were in close proximity and aligned with the required orientations for stable association. Existing studies in literature have assumed that *trans*-splicing is more efficient in the cytosol due to greater mobility of the precursors for achieving the required orientations to associate with each other [228,388]. It is unknown whether *trans*-splicing could also occur once the precursors have localised to the membranes. The efficiency of *trans*-splicing in membranes is likely to be much lower as there are more stringent requirements: both

precursors must insert into the membrane with the same orientation and their association would depend on lateral diffusion within the membrane bilayer, which is influenced by the composition of the membrane. For the case of pSRII, the low *trans*-splicing efficiency could potentially suggest fast membrane localisation of the precursors, despite the lack of signal peptide in both precursors. Solubility enhancement tags can be added to both precursors in future studies to improve the expression yield and ligation efficiency, as has been previously reported for proteorhodopsin [228]. Careful assessment will be required to ensure that addition of such tags does not interfere with the structure of each precursor. Considering that most α-helical membrane proteins are co-translationally inserted into the membrane [142,144], delaying membrane localisation of the precursors could result in protein misfolding and/or partitioning to inclusion bodies. Co-expression of chaperones, such as GroEL-GroES, might help to avoid misfolding and improve proper folding of the precursors [400,421]. In addition, the effects of the *Npu* DnaE intein fusion on the rate and orientation of membrane insertion of the pSRII fragments are unknown. Future studies should include control experiments where the reconstitution of the native chromophore is attempted by co-expressing and co-purifying the N-terminal and the C-terminal fragment of pSRII in the absence of the intein fusion. Different expression vectors could also be tested to explore whether the expression level of each pSRII fragment has any effects on the rate of membrane insertion, protein folding, and reconstitution of the native chromophore.

The *trans*-splicing reaction might also have been hampered by unsuccessful liberation of the C-intein from the branched thioester intermediate, as evidenced by the extra band running between 30 and 40 kDa observed in fraction E purified from dual-expression in C41(DE3) cells. The C-intein is cleaved from the branched intermediate upon cyclisation of Asn, the last residue of the C-intein. The reason behind unsuccessful Asn cyclisation is unclear. Given the importance of the +2 residue in the C-extein in positioning His-125, the general acid/base of the internal cyclisation reaction, future studies would involve exploring alternative sequences at the splice junction. In addition, the native *Npu* DnaE extein sequence, AEY/CFN, must also be investigated as a control to ascertain whether the AEK/CWL splice junction sequence is suitable for *trans*-splicing. Further optimisation of the splice junction sequence could also decrease the amount of C-terminal cleavage or other side reactions.

The location of the split site may also require careful evaluation in future studies. Although the insertion of four residues appeared to be inert in the full-length pSRII-KCWL mutant, introducing a split site between the two β-strands of the hairpin may interfere with the formation of the hairpin structure. Since the H-bonds between Arg-66 and residues in helix D might be important in anchoring the β-hairpin, it is proposed that an alternative split site in the BC loop could be introduced before the hairpin.

One approach for determining the rate limiting factors of *in vivo trans*-splicing is attempting *trans*-splicing *in vitro* in the presence of defined reaction conditions and molar ratios of the two precursors [400,422]. However, this approach will not be trivial since the N-precursor was DDM-insoluble. *In vitro* studies would therefore require the screening of different detergents and buffer conditions in order to identify conditions which are suitable for solubilising both precursors and for functional folding of the ligated pSRII-KCWL.

In summary, the studies presented in this chapter lay considerable foundation for *in vivo* assembly of pSRII using split inteins. Further optimisation likely requires exploring alternative extein sequences at the splice site, or selecting another location in the BC loop as the split site. The overarching goal of establishing these proof-of-concept studies is to examine how different loops and helices affect the folding pathway of pSRII. It is envisioned that systematic comparison of experimental results will be difficult for *in vivo trans*-splicing across different loops of pSRII, because of the high likelihood of resorting to different –3 to +3 extein sequences at the splice sites in different loops, and the large number of factors which contribute towards the stability of precursors, the yield of the ligated product, and the rate of protein *trans*-splicing. Future studies on assembling pSRII from complementary fragments will likely require a combination of *in vitro* and *in vivo* approaches.

6 Concluding remarks and future directions

Considerable progress has been made towards understanding the folding process of membrane proteins by experimental and computational methods over the past few decades. The challenging nature of membrane protein folding studies remains evident from the small number of such studies available in literature compared to soluble protein folding studies.

The blueprints for proteins to obtain their native folds are, in principle, embedded in their sequences. Existing studies on bacteriorhodopsin and bovine rhodopsin have led to the proposal of the two-stage model and the long-range interactions model for α-helical membrane protein folding. These studies raise intriguing questions which serve as the driving questions of this thesis – Do structurally homologous membrane proteins have the same folding mechanism? What factors underpin different folding mechanisms being adopted by membrane proteins?

Chapters 2 and 3 have provided detailed insights on the unfolding and refolding mechanisms of pSRII, an archaeal member of the rhodopsin family. The SDS-denatured, acid-denatured and apo-sensory opsin states were characterised using biophysical techniques and solution-state NMR. pSRII remained largely helical when unfolded, and tertiary structure was disrupted. The loss of native retinal-protein contacts led to changes in backbone amide dynamics, with increased amount of intermediate timescale exchange observed across different parts of the protein depending on the unfolding condition employed. There are very few solution-state NMR studies on the denatured states of full-length membrane proteins. Future studies will also require experimental techniques which provide sufficient structural and time resolution particularly for characterising motions in the μs–ms timescale, which are important in helix-helix interactions. Kinetic studies on the unfolding and refolding of pSRII in SDS/diheptanoylphosphatidylcholine (c7-DHPC) mixed detergent micelles demonstrated pSRII as one of the very few membrane proteins which refolds from SDS-denatured states via a non-two-state mechanism, thus distinguishing pSRII from the two-state folder, bacteriorhodopsin, and bovine rhodopsin, for which the SDS-denatured state has evaded

in vitro refolding in existing literature. This is the first demonstration of a heptahelical transmembrane protein folding via a complex non-two-state mechanism.

The detailed molecular insights on pSRII illustrate the diversity of folding mechanisms across the rhodopsin family and α-helical membrane proteins. Existing studies in literature and the results presented in this thesis have demonstrated that studying the native fold alone is insufficient to account for the diverse folding mechanisms across the rhodopsin family. In the soluble protein folding field, many experimental studies have been performed to compare the folding pathways of proteins with similar tertiary structures but divergent sequences [423,424]. Quantitative comparisons of folding transition state structures, determined by ϕ-value analysis of folding kinetics, show that certain proteins, such as the SH3 domains and Ig-like domains, share common transition state structures within the families, whereas other proteins, such as Suc1 and Cks1, possess rather different transition state structures despite high sequence and structural similarities. Specific native tertiary structures impose different constraints on the choice of folding pathways, yielding a continuum of different extents of transition state structure conservation. It has been theorised that the transition states of proteins with diverse folding pathways are influenced more by variations in protein sequence and local structural propensity, thus leading to differences in how the native tertiary structure is formed [423,424].

Detailed studies on transition states have thus far been sparse for membrane proteins due to challenges in identifying conditions which permit reversible unfolding *in vitro*. Having demonstrated reversible unfolding of pSRII in Chapter 2, future experiments should include ϕ-value analysis to compare the transition state structures of pSRII against those of other rhodopsin family members. Extending the summarised findings from the soluble protein field to membrane protein folding studies, the caveat with direct comparison of folding pathways across the rhodopsin family is that although all rhodopsin family members are composed of seven transmembrane helices with a covalently-bound retinal chromophore (i.e. similar topologies), there are notable differences in interhelical packing and loop structures (i.e. differences in tertiary structure) between microbial and metazoan rhodopsins. These structural differences may already partially account for different unfolding and refolding mechanisms elucidated by *in vitro* studies. In addition, there is no detectable sequence

homology between microbial and metazoan rhodopsins, although there is clear sequence identity within the respectable groups [425]. Future studies on the folding mechanisms of G protein-coupled receptors (GPCRs), which have the same topology and highly similar tertiary structure to metazoan rhodopsins, might yield better insights on how sequence variation and native tertiary structure contribute towards the folding pathway respectively. In fact, a recent study on the refolding of the β_1-adrenergic receptor (β_1AR) on a solid support shows that the protein cannot be refolded from the SDS-denatured state whereas urea denaturation is reversible, indicating different responses to urea but similar irreversibility of SDS-denaturation compared to bovine rhodopsin [109]. It is hoped that exploring the folding mechanisms of more heptahelical membrane proteins will ultimately shed insights as to which aspects of the folding mechanism (e.g. folding nucleation sites, transition states or roles of particular residues) are conserved between structurally homologous proteins and between evolutionarily related proteins.

As discussed in Chapters 2 and 3, comparable membrane mimetics and experimental conditions should ideally be used in comparative studies of folding mechanisms across different membrane proteins. A set of experimental conditions has been recommended to standardise and facilitate comparison of soluble protein folding studies [426]. Although similar guidelines will likewise be highly valuable for comparative studies of membrane protein folding, identifying conditions which permit reversible unfolding/refolding studies for any particular membrane protein is already a laborious trial-and-error process of finding the suitable membrane mimetics and buffer conditions, let alone proposing standardised conditions which are applicable across multiple membrane proteins [18]. It is therefore paramount to further study the molecular properties of different mimetics and the determinants of solubility and functionality of membrane proteins [427]. For example, a recent study has explored the effects of different membrane mimetics on urea-mediated unfolding of LeuT, including different detergent micelles, liposomes and lipid bilayers [110].

H-bonding is one of the major stabilising forces for protein structure. Accurate descriptions of transient H-bonds in denatured states and the processes of H-bond formation during protein folding will provide valuable information about the folding pathway. In Chapter 4, $^1J_{NH}$ measurements were performed to assess H-bonding in folded and denatured membrane

proteins, albeit with somewhat limited success due to challenges with data precision and/or accuracy attributable to cross-correlated relaxation. Utilisation of perdeuterated samples and further optimisation of pulse sequences will provide more conclusive assessment of the applicability of $^{1}J_{NH}$ measurements towards studying H-bonding and obtaining detailed molecular insights on the structures of transition states and folding intermediates.

One of the key differentiators of the two-stage model vs. the long-range interactions model is the contribution of loop regions towards membrane protein folding. Analysis of the crystal structures of 41 α-helical membrane proteins shows significant variations in the extramembranous regions [428]. 53% of the loops are found to be stretched, and impose geometrical constraints on the distances between the two connected helices. Non-stretched loops tend to prefer conformations which keep hydrophobic residues buried. Such conformational preferences may be extended into the transmembrane regions, thus affecting the assembly of surrounding helices and not just the consecutive ones. Such hydrophobic interactions therefore constitute as long-range interactions. For 70% of the membrane proteins investigated, at least one side is significantly affected by such long-range interactions driven by the compacting of longer loops [428].

Experimental assessment of the roles of loops in protein folding has traditionally been performed by assembling different complementary fragments with the protein being split at different loop regions. In Chapter 5, attempts were made, albeit unsuccessfully, to assemble pSRII using split intein. Improving the yield of the ligated product would require careful choice of protein sequences constituting the splice junction and possible utilisation of solubilisation enhancement tags. In future studies, it would be valuable to study how protein sequence and local structure propensities influence the contribution of different loops towards short- and long-range interactions in the contexts of protein folding and stability.

It must be emphasised that the biological context of protein folding is much more complex than the *in vitro* experimental systems which are currently employed in soluble and membrane protein folding studies. Additional factors include *in vivo* folding rates, co-translational folding and insertion into the membrane, interactions with chaperones and other protein quality control mechanisms, and subcellular localisation. In recent years, novel

approaches have been developed for studying the folding and membrane insertion of ribosome nascent chains during *in vitro* cell-free expression [429] and from structural studies on purified ribosome-SecY protein complexes with nascent membrane protein chains [430]. These approaches hold great promise for obtaining structural and mechanistic insights on membrane protein folding and the roles of different biological factors.

In summary, the current opinion is that "the devil is in the details" when it comes to understanding how membrane proteins fold. It is anticipated that combined *in vitro*, *in silico* and *in vivo* approaches will continue to play instrumental roles in elucidating the complex intricacies of membrane protein folding.

Appendices

Appendix

A Materials and methods

A.1 Media and Buffers

A.1.1 Molecular biology

Medium	Composition
TF1	10 mM MOPS (pH 6.8), 10 mM $CaCl_2$
TF2	10 mM MOPS (pH 6.8), 75 mM $CaCl_2$
TF3	50 mM MOPS (pH 7.5), 75 mM $CaCl_2$, 5% glycerol, 1 mM EDTA, 100 mM KCl
LB	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl
LB Agar	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 17.5 g/L bacto-agar
2TY	16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl
M9 (pSRII) [216]	47.7 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 4 µM $ZnSO_4$, 1 µM $MnCl_2$, 0.7 µM H_3BO_3, 0.7 µM $CuSO_4$, 2 µM $FeCl_3$, 5 µM $MgSO_4$, 100 µM $CaCl_2$, 0.5% glucose, 0.1% $^{15}NH_4Cl$, 50 mg/L kanamycin, 34 mg/L chloramphenicol
M9 (OmpX)	47.7 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 1 µM $FeCl_3$, 2 mM $MgSO_4$, 100 µM $CaCl_2$, 50 µM $ZnSO_4$, 1 mg/L biotin, 1 mg/L choline chloride, 1 mg/L folic acid, 1 mg/L niacinamide, 1 mg/L D-pantothenate, 1 mg/L pyridoxal-HCl, 0.1 mg/L (-)riboflavin, 5 mg/L thiamine, 0.6% glucose, 0.1% $^{15}NH_4Cl$, 30 mg/L kanamycin

Table A.1: Composition of media and buffers for molecular biology

A.1.2 Biochemistry

Buffer	Composition
Native loading dye (2×)	62.5 mM Tris-HCl (pH 6.8), 25% glycerol, 0.01% bromophenol blue
RIPA	50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 0.1% SDS, 1% Triton X-100, 0.5% sodium deoxycholate, 1 mM TCEP-HCl
NuPAGE® transfer buffer	25 mM Bicine, 25 mM bis-tris, 1 mM EDTA, 0.05 mM chlorobutanol, 10% (one blot) or 20% (two blots) methanol
Tris-buffered saline (TBS)	25 mM Tris-HCl (pH 7.2), 150 mM NaCl
Tris-buffered saline with Tween-20 (TBST)	25 mM Tris-HCl (pH 7.2), 150 mM NaCl, 0.05% Tween-20

Table A.2: Composition of buffers for biochemistry techniques

A.1.3 Protein purification

Protein	Buffer	Composition
pSRII	Buffer A	50 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$
	Buffer S	50 mM MES-NaOH (pH 6.5), 300 mM NaCl, 5 mM imidazole
	Buffer W1	50 mM MES-NaOH (pH 6.5), 300 mM NaCl, 25 mM imidazole
	Buffer W2	50 mM MES-NaOH (pH 6.5), 300 mM NaCl, 50 mM imidazole
	Buffer E	50 mM Tris-HCl (pH 7.0), 300 mM NaCl, 150 mM imidazole
	Buffer E300	50 mM Tris-HCl (pH 7.0), 300 mM NaCl, 300 mM imidazole
	Buffer N7	20 mM NaH_2PO_4/Na_2HPO_4 (pH 7.0), 50 mM NaCl, 0.05% NaN_3
	Buffer N6	50 mM NaH_2PO_4/Na_2HPO_4 (pH 6.0), 50 mM NaCl, 0.05% NaN_3
OmpX	Buffer TE	20 mM Tris-HCl (pH 8.0), 5 mM EDTA
	Buffer G	50 mM Tris-HCl (pH 8.5), 100 mM NaCl, 5 mM EDTA, 6 M GdnHCl
	Buffer U	20 mM Tris-HCl (pH 8.5), 5 mM EDTA, 8 M Urea
	Buffer R	50 mM Tris-HCl (pH 8.5), 5 mM EDTA, 0.5 M L-arginine-HCl, 0.5% DPC
	Buffer D	20 mM Tris-HCl pH (8.5), 5 mM EDTA, 100 mM NaCl
	Buffer N	20 mM NaH_2PO_4/Na_2HPO_4 (pH 6.8), 100 mM NaCl, 5 mM EDTA
	Buffer X	20 mM NaH_2PO_4/Na_2HPO_4 (pH 6.5), 8 M urea, 0.1 mM NaN_3

Table A.3: Compositions of buffers used for protein purification
MES = 2-(N-morpholino)ethanesulphonic acid
Tris = Tris(hydroxymethyl)aminomethane

A.2 Molecular Biology

A.2.1 Sensory rhodopsin II (pSRII)

The wild type full-length sensory rhodopsin II *psopII* gene (EMBL-EBI accession number CAI50508) from *Natronomonas pharaonis* was previously cloned between the NcoI and XhoI sites of the pET-28b(+) vector (Novagen), which adds the amino acid sequence LEHHHHHH to the C-terminus and also carries a kanamycin resistance gene. The clone was obtained from Prof. Javier Navarro (University of Texas Medical Branch, Galveston, USA). pSRII expression is under the control of a T7 promoter. The final construct encodes residues 1–241 of *Natronomonas pharaonis* pSRII plus the C-terminal hexahistidine (His_6) tag:

```
MVGLTTLFWL GAIGMLVGTL AFAWAGRDAG SGERRYYVTL VGISGIAAVA  50
YAVMALGVGW VPVAERTVFV PRYIDWILTT PLIVYFLGLL AGLDSREFGI 100
VITLNTVVML AGFAGAMVPG IERYALFGMG AVAFIGLVYY LVGPMTESAS 150
QRSSGIKSLY VRLRNLTVVL WAIYPFIWLL GPPGVALLTP TVDVALIVYL 200
DLVTKVGFGF IALDAAATLR AEHGESLAGV DTDTPAVADL EHHHHHH     247
```

A.2.2 Outer membrane protein X (OmpX)

The plasmid used for expressing OmpX was a gift from Prof. Sebastian Hiller (Biozentrum, University of Basel). The plasmid was constructed by inserting the N-terminally truncated *OmpX* gene from *E. coli* strain K-12 (Uniprot P0A917), encoding for amino acids 24-171 of OmpX, into a pET-28b(+) vector (Novagen) containing a kanamycin resistance gene. OmpX expression is under the control of a T7 promoter.

The encoded sequence for Met-OmpX(24–171) is:

```
MATSTVTGGY AQSDAQGQMN KMGGFNLKYR YEEDNSPLGV IGSFTYTEKS  50
RTASSGDYNK NQYYGITAGP AYRINDWASI YGVVGVGYGK FQTTEYPTYK 100
HDTSDYGFSY GAGLQFNPME NVALDFSYEQ SRIRSVDVGT WIAGVGYRF  149
```

A.2.3 Cloning pSRII^{N}(KCWL)-DnaE^{N} and DnaE^{C}-pSRII^{C}(KCWL)

The pSKDuet01 (Addgene plasmid #12172) and pSKBAD2 (Addgene plasmid #15335) plasmids [387] were gifts from Dr. Hideo Iwai (University of Helsinki, Finland). Amplification of whole vectors and insert sequences by polymerase chain reaction (PCR) were performed with Phusion DNA polymerase (New England BioLabs). The reaction mix (prepared on ice) (Table A.4) and thermocycle parameters (Table A.5) were defined according to manufacturer's instructions. Sequences of the DNA primers, annealing temperature and extention time are defined in Table A.7.

Component	Volume (µl)	Final Concentration
Nuclease-free water	30–34	
5× Phusion HF buffer	10.0	1×
10 mM dNTPs	1.0	200 µM
10 µM Forward primer	2.5	0.5 µM
10 µM Reverse primer	0 or 2.5	0.5 µM
Template DNA	0.5–2.0	10 ng plasmid DNA
DMSO	1.5	3%
Phusion DNA polymerase	0.5	1 unit
Total Volume	**50**	

Table A.4: Composition of PCR reaction mix

Step	Temperature (°C)	Time
Initial denaturation	98	30 s
30 cycles	98 45–73 72	10 s 30 s 30 s per kb
Final extension	72	10 min
Hold	4-10	∞

Table A.5: Thermocycle parameters

pSRII^{N} and pSRII^{C} were first inserted into pSKDuet01 and pSKBAD2 by ligation independent cloning. The insert sequences were PCR-amplified from pET28b(+)-pSRII-His$_6$. The vectors were linearised by PCR. The appropriate pairs of PCR-amplified inserts and vectors (boxed pairs in Table A.7) were then assembled by Gibson assembly (New England BioLabs) according to the manufacturer's instructions. The reaction mix (Table A.6) was prepared on ice, and the mixture was incubated for 1 h at 50 ºC.

Component	Volume
Insert	3–5 µl unpurified PCR product
Vector	1 µl unpurified PCR product
Gibson assembly master mix (2×)	10 µl
Deionised H_2O	4–6 µl
Total volume	**20 µl**

Table A.6: Gibson assembly reaction mix

2–4 µl of the Gibson assembly reaction mix was then transformed into competent *E. coli* DH5α cells. DNA sequence was confirmed by Sanger DNA sequencing (DNA Sequencing Facility, Department of Biochemistry, University of Cambridge).

Strep-tag II insertion before pSRII^{C}, insertion of the appropriate sequences for the splice site, and insertion of a flexible linker sequence (Ser-Ala) between the start codon and Strep-tag II before pSRII^{C}, were all achieved by single primer PCR. Strep-tag II insertion after pSRII^{N} was achieved by encoding the Strep-tag II as the 5' overhangs of the forward and reverse

primers. 2 µl of each of these PCR products were transformed directly into *E. coli* DH5α cells as un-ligated plasmids, and the sequence confirmed by Sanger DNA sequencing.

The final constructs encoded the following sequences:

pET28b(+)-pSRII^{N}(KCWL)-Int^{N}-Strep:

```
MVGLTTLFWL  GAIGMLVGTL  AFAWAGRDAG  SGERRYYVTL  VGISGIAAVA   50
YAVMALGVGW  VPVAEKCLSY  ETEILTVEYG  LLPIGKIVEK  RIECTVYSVD  100
NNGNIYTQPV  AQWHDRGEQE  VFEYCLEDGS  LIRATKDHKF  MTVDGQMLPI  150
DEIFERELDL  MRVDNLPNSA  WSHPQFEK                           178
```

pBad-Strep-Int^{C}-pSRII^{C}(KCWL)-His$_6$:

```
MSAWSHPQFE  KSAMIKIATR  KYLGKQNVYD  IGVERDHNFA  LKNGFIASNC   50
WLRTVFVPRY  IDWILTTPLI  VYFLGLLAGL  DSREFGIVIT  LNTVVMLAGF  100
AGAMVPGIER  YALFGMGAVA  FIGLVYYLVG  PMTESASQRS  SGIKSLYVRL  150
RNLTVVLWAI  YPFIWLLGPP  GVALLTPTVD  VALIVYLDLV  TKVGFGFIAL  200
DAAATLRAEH  GESLAGVDTD  TPAVADLEHH  HHHH                   232
```

A.2.4 Insertional mutagenesis

Insertion of the residues Lys-Cys-Trp-Leu (KCWL) into pET28b(+)-pSRII-His$_6$ was achieved by single primer PCR. 2 µl PCR products were transformed directly into *E. coli* DH5α cells as un-ligated plasmids, and the sequence confirmed by Sanger DNA sequencing. The final construct encodes:

```
MVGLTTLFWL  GAIGMLVGTL  AFAWAGRDAG  SGERRYYVTL  VGISGIAAVA   50
YAVMALGVGW  VPVAEKCWLR  TVFVPRYIDW  ILTTPLIVYF  LGLLAGLDSR  100
EFGIVITLNT  VVMLAGFAGA  MVPGIERYAL  FGMGAVAFIG  LVYYLVGPMT  150
ESASQRSSGI  KSLYVRLRNL  TVVLWAIYPF  IWLLGPPGVA  LLTPTVDVAL  200
IVYLDLVTKV  GFGFIALDAA  ATLRAEHGES  LAGVDTDTPA  VADLEHHHHHH  251
```

(*Next page*)

Table A.7: Primers

Pairs of insert and vector used in Gibson Assembly reactions are boxed. Other PCR products were transformed into *E. coli* DH5α cells as un-ligated plasmids. Regions of primer sequences which anneal to their respective templates are shown in red font.

Template	Purpose	Primers (5'→3')		Annealing temp (°C)	Extension time (s)
pET28b(+)-pSRII^{N}(KCWL)-Int^{N}-Strep					
pET28b(+)-pSRII-His$_6$	Insert amplification	For	GGAGATATACCATGGTGGGACTTACGACCC	65	17
		Rev	AACAGGATCCCACGGGAACCCAGCCGAC		
pSKDuet-Int^{N} (pSKDuet01)	Vector amplification	For	CCCGTGGGATCCTGTTTAAGCTATG	55	123
		Rev	CCCACCATGTATATCTCCTTATTAAAGTTAAACAAAATTATTTC		
pSKDuet-pSRII^{N}-Int^{N}	Strep-tag II insertion	For	CAGTTTGAAAAATAAGCTTGCGGCCGCATAATG	69	129
		Rev	CGGATGGCTCCAGGCAGAATTCGGCAAATTATCAACCCGC		
pSKDuet-pSRII^{N}-Int^{N}-Strep	KCWL splice site insertion	-	CTGGGTTCCCGTGGCCGAAAAGTGTTTAAGCTATGAAACGGAAATATTGACAG	64	129
pSKDuet-pSRII^{N}(KCWL)-Int^{N}-Strep	Insert amplification	For	AAGAAGGAGATATACCATGGTGGGACTTACG	54	17
		Rev	CAGCCGGATCTTATTTTTCAAACTGCG		
pET28b(+)-pSRII-His$_6$	Vector amplification	For	GTTTGAAAAATAAGATCCGGCTGCTAAC	58	157
		Rev	TCCCACCATGGTATATCTCCTTCTTAAAGTTAAAC		
pBad-Strep-Int^{C}-pSRII^{C}(KCWL)-His$_6$					
pET28b(+)-pSRII-His$_6$	Insert amplification	For	GTTTCAATGGAACTGTTTTCGTCCCCCGGTAC	73	17
		Rev	CTTCTCTCAGTGGTGGTGGTGGTGGTGCTCGAG		
pBad-Int^{C} (pSKBAD2)	Vector amplification	For	CCACCACTGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAG	72	141
		Rev	CAGTTCCATTGAAACAATTAGAAGCTATGAAGCCATTTTTGAGTGC		
pBad-Int^{C}-pSRII^{C}-His$_6$	Strep-tag II insertion	-	GAAGGAGATATACATATGTGGAGCCATCCGCAGTTTGAAAAATCTGCCATGATCAAAATAGCCACACGTAAATATTTAG	64	141
pBad-Strep-Int^{C}-pSRII^{C}-His$_6$	KCWL splice site insertion	-	ATAGCTTCTAATTGTTGGCTCCGGACTGTTTTCGTCCCCCG	64	141
pBad-Strep-Int^{C}-pSRII^{C}(KCWL)-His$_6$	Linker insertion	-	GATATACATATGTCTGCCTGGAGCCATCCGC	70	141
pET28b(+)-pSRII-KCWL-His$_6$					
pET28b(+)-pSRII-His$_6$	Insert KCWL	-	GTTCCCGTGGCCGAAAAGTGTTGGCTCCGGACTGTTTTCGTC	52	179

A.2.5 Transformation

Glycerol stocks of *E. coli* Tuner(DE3)LacI cells (Novagen) and BL21(DE3) cells (Novagen) were made competent using the $CaCl_2$ method [431]: 0.2 ml of an overnight culture was added to 10 ml 2TY and grown at 37 ºC until OD_{600} = 0.4–0.6. Cells were pelleted (4000 × *g*, 10 min, 4 ºC), resuspended in 7 ml ice-cold 0.1 M $CaCl_2$, kept on ice for 1 h, pelleted again, resuspended in 0.5 ml ice-cold 0.1 M $CaCl_2$, and incubated with ~100 ng DNA on ice for 30 min. The cells were heat-shocked (42 °C for 90 s, then on ice for 3 min). 0.7 ml LB was added and the cells were incubated at 37 °C for 45 min. Transformed cells were grown overnight at 37 °C on LB agar with the appropriate antibiotics.

E. coli C41(DE3) cells were a gift from George Sophocleous. 0.2 ml of an overnight culture of C41(DE3) cells was added to 10 ml 2TY and grown at 37 ºC until OD_{600} = 0.4–0.6. Cells were pelleted (4000 × *g*, 10 min, 4 ºC), resuspended in 10 ml TF1, pelleted again, resuspended in 1 ml TF2, and kept on ice for 16 h. 1 µl (~100 ng) DNA was added to 100 µl of super-competent cells and 99 µl TF3. The mixture was incubated on ice for 45 min. The cells were heat-shocked (42 °C for 10 min, then on ice for 3 min). 0.7 ml LB was added and the cells were incubated at 37 °C for 45 min. Transformed cells were grown overnight at 37 °C on LB agar with the appropriate antibiotics.

A.3 Biochemistry

A.3.1 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was carried out according to Laemmli [432]. Gels were poured in 1.0 mm mini-gel casettes (NC2010, Thermo Fisher Scientific) with the components of the stacking gel and the resolving gel detailed in Table A.8.

Protein samples were mixed with the appropriate volume of 4× NuPAGE® LDS sample buffer (Thermo Fisher Scientific) for denaturing SDS-PAGE. Alternatively, 2× native loading dye [270] was used for the analysis of aggregation states in unfolded protein samples. SDS-PAGE analysis was performed on a 12% polyacrylamide gel (Table A.8) at 90 V for 15 min followed by 180 V for 40 min in NuPAGE® MES SDS running buffer (Thermo Fisher

Scientific). PageRuler prestained protein ladder (10–180 kDa) (Thermo Fisher Scientific) was used as molecular weight markers. Protein bands were visualised using InstantBlue (Expedeon) or the Pierce™ Silver Stain Kit (Thermo Fisher Scientific) according to manufacturers' instructions.

Ingredient	12% Resolving gel	6% Stacking gel
milliQ-H_2O	1.69 ml	1 ml
40% acrylamide	1.75 ml	0.3 ml
2% bis-acrylamide	0.93 ml	0.15 ml
0.5 M Tris-HCl (pH 6.8)	-	0.5 ml
1.5 M Tris-HCl (pH 8.8)	1.5 ml	-
10% SDS	60 µl	20 µl
10% APS	60 µl	20 µl
TEMED	6 µl	2 µl

Table A.8: Compositions of stacking gel and resolving gel for SDS-PAGE
SDS: electrophoresis grade sodium dodecyl sulphate (Sigma-Aldrich); APS: ammonium persulphate; TEMED: tetramethylethylenediamine

A.3.2 Western blot

Proteins separated by SDS-PAGE were transferred onto Immobilon®-P (PVDF) membranes with 0.45 µm pore size (Merck Millipore) according to protocol from Invitrogen, using NuPAGE® transfer buffer, at a constant voltage of 30 V for 1 h at room temperature. Membranes were blocked with 2.5% bovine serum albumin (BSA) in TBST buffer with gentle agitation for 1 h at room temperature or overnight at 4 °C. 50 µg/ml avidin was also added to the blocking solution to minimise detection of endogenous biotin carrier proteins during Strep-tag II detection. For His_6-tag detection, membranes were washed for 2 × 10 min with TBST, incubated in 1:5000 dilution of HisProbe™-HRP (horseradish peroxidase) conjugate (Thermo Fisher Scientific) for 1 h at room temperature, and washed for 4 × 10 min with TBST. For Strep-tag II detection, membranes were washed for 3 × 5 min with TBST, incubated in 1:10000 dilution of Precision Protein™ StrepTactin-HRP Conjugate (Bio-Rad Laboratories) for 1 h at room temperature, and washed for 2 × 5 min with TBST and once for 5 min in TBS. Signals were developed using Pierce™ ECL Western Blotting substrate

(Thermo Fisher Scientific) according to the manufacturer's instructions, and detected by film (Fujifilm, Japan).

A.4 Protein expression and purification

A.4.1 pSRII

Expression of pSRII in *E. coli* Tuner(DE3)LacI cells (Novagen) was performed as described by Gautier *et al* [216]. Single transformed colonies from an LB agar plate were inoculated into 50 ml LB media with 34 μg/ml chloramphenicol and 50 μg/ml kanamycin, and incubated overnight (37 °C, 200 rpm). Each overnight culture was then transferred to 500 ml LB media to give starting OD_{600} = 0.1. Cells were grown at 37 °C with shaking at 200 rpm until OD_{600} = 0.8–1.0. Expression of pSRII was induced by adding 1 mM IPTG. The cells were further incubated for 10 h at 25 °C, 200 rpm, with 10 μM all-*trans* retinal (Sigma-Aldrich; dissolved in absolute ethanol and kept at -20 °C in the dark) supplemented at 2 h-intervals. Cells were centrifuged (5000 × *g*, 20 min, 4 °C) and the cell pellets were frozen (–20 °C).

For the expression of uniformly ^{15}N-labelled pSRII, M9 media was used for overnight starter cultures and for protein expression (Table A.1). For the expression of ^{19}F-5-Trp-labelled pSRII, 60 mg L^{-1} 5-fluoroindole (Sigma-Aldrich; dissolved in sterile DMSO) [334] was added to un-induced cultures at $OD_{600} \approx 0.8$. The cultures were further incubated at 37 °C, 200 rpm for 10 min before pSRII expression was induced by adding 1 mM IPTG and 10 μM all-*trans* retinal was supplemented at 2 h-intervals for 10 h, as described for the expression of unlabelled pSRII.

The cells were resuspended in 1/40th of the original culture volume in ice-cold lysis Buffer A with protease inhibitor cocktail (Sigma-Aldrich, P2714). The lysate was passed thrice through an Emulsiflex-C5 High Pressure Homogeniser (Avestin, Inc.) operating at 50 Psi. Crude membranes were pelleted by ultracentrifugation (100,000 × *g*, 90 min, 4 °C) and resuspended in 1/75th of the original culture volume in Buffer S with protease inhibitor cocktail (Sigma-Aldrich). The suspension was homogenised on ice (T18 basic ULTRA-TURRAX®, IKA®). 1.5% n-dodecyl-β-D-maltoside (DDM) (Melford) was added

and the solution solubilised by rotation (in the dark, 16 h, 4 °C). Insoluble material was removed by ultracentrifugation (100,000 × *g*, 60 min, 4 °C).

Ni-NTA agarose beads (Qiagen) (4 ml beads per litre of culture) were equilibrated with 5 column volumes (CV) milliQ water, followed by 5 CV Buffer S. The beads were mixed with the supernatant containing solubilised membrane proteins and lipids by rotation (in the dark, 2.5 h, 4 °C). Beads were filtered out using gravity-flow columns (BioRad Econo-Pac columns) and washed with 25 CV Buffer S, 25 CV Buffer W1 and 7.5 CV Buffer W2. All wash buffers were ice-cold and contained 0.06% DDM.

The protein was eluted in 2.5 CV Buffer E, followed by 2 CV Buffer E300, both containing with 0.1% DDM. The sample was concentrated to 0.5–1 ml using 50 kDa cut-off centrifugal concentrators (Sartorius) and heated to precipitate impurities (55 °C, 60 h). The sample was then resuspended in Buffer N7 + 0.03% DDM, and centrifuged to pellet precipitated impurities.

The protein was mixed by rotation (in the dark, 2.5 h, 4 °C) with Ni-NTA beads that were pre-equilibrated in Buffer N7. The beads were filtered out using gravity-flow columns, and washed with 50 CV Buffer N7 to remove DDM. 2 CV Buffer N7 was retained in each column and 0.6% 1,2-diheptanoyl-sn-glycero-3-phosphocholine (c7-DHPC) (Avanti Polar Lipids or Anatrace) was added before the sample was mixed by rotation (in the dark, 16 h, 4 °C). The flow-through was discarded and the column was washed with 50 CV Buffer N7 with 0.1% c7-DHPC. The protein was eluted with 2.5 CV Buffer N6 + 0.1% c7-DHPC + 300 mM Imidazole. Imidazole was removed from the eluted protein by at least 3 rounds of concentration/dilution in a centrifugal concentrator (50 kDa cut-off) using 60 ml Buffer N6 + 0.1% c7-DHPC. The final sample was concentrated to ~500 µl. Additional c7-DHPC (from a 20% c7-DHPC stock solution) was added to the sample to achieve a final c7-DHPC concentration of ~2%.

Expression and purification of pSRII-KCWL were performed as described for WT-pSRII. Size exclusion chromatography was done using a Superdex 200 Increase 10/300 GL column

(GE Healthcare) equilibrated in Buffer N6 + 0.05% DDM and running at 0.5 ml/min at 4 °C. Each eluted fraction was 1 ml.

A.4.2 pSRII^{N}(KCWL)-Int^{N}-Strep and Strep-Int^{C}-pSRII^{C}(KCWL)-His$_6$

Different expression temperatures, order of induction, and inducer concentrations were screened, as detailed in Chapter 5. For test expressions, single transformed colonies from an LB agar plate were inoculated into 10 ml LB media with the appropriate antibiotics, and incubated overnight (37 °C, 200 rpm). Each overnight culture was then transferred to 80 ml LB media to give starting OD_{600} = 0.1. Cells were grown at 37 °C with shaking at 200 rpm until OD_{600} = 0.8–1.0. Expression of pSRII^{N}(KCWL)-Int^{N}-Strep or Strep-Int^{C}-pSRII^{C}(KCWL)-His$_6$ were induced by addition of IPTG or L-arabinose, respectively. 5 ml aliquots of cell culture which were taken at specific time-points, and the final ~40 ml cultures were centrifuged (4000 × *g*, 10 min, 4 °C). All cell pellets were kept frozen (–20 °C) before being used for Western blotting.

To obtain whole cell lysates for Western blotting, the cell pellets were resuspended in ice-cold RIPA buffer and kept on ice for 1 h. The amount of RIPA buffer used for each sample was based on the OD_{600} measured by UV/vis spectroscopy and cell density was adjusted to the equivalent of a final OD_{600} = 5.0. The lysed cells were mixed with 4× NuPAGE® LDS sample buffer (Thermo Fisher Scientific) and analysed by SDS-PAGE and Western blotting.

The pellets from the final ~40 ml *E. coli* cultures were used for determining the cellular localisation of the expressed proteins. Each pellet was resuspended in 5–10 ml RIPA buffer, with the volume adjusted to yield comparable cell density across samples. The suspension was left on ice for 1 h. Unlysed cells were pelleted by centrifugation at low speed (1000 × *g*, 4 ℃, 10 min). The supernatant was ultracentrifuged (100,000 × *g*, 4 ℃, 90 min) to pellet the crude membranes. The resulting pellet was resuspended and incubated with rotation in the original volume of RIPA buffer + 1.5% DDM for 30 min at room temperature. The samples were mixed with 4× NuPAGE® LDS sample buffer (Thermo Fisher Scientific) and analysed by SDS-PAGE and Western blotting.

For large-scale (3 L) co-expression in *E. coli* Tuner(DE3)LacI or C41(DE3) cells, single transformed colonies from an LB agar plate were inoculated into 50 ml LB media with 100 μg/ml ampicillin and 50 μg/ml kanamycin (plus 34 μg/ml chloramphenicol for Tuner(DE3)LacI cells), and incubated overnight (37 °C, 200 rpm). Each 50 ml culture was then transferred to 500 ml LB media to give starting OD_{600} = 0.1. Cells were grown at 37 °C with shaking at 200 rpm until OD_{600} = 0.8–1.0. Expression of Strep-Int^C-$pSRII^C$(KCWL)-His_6 was first induced by adding 0.05% L-arabinose, and 10 μM all-*trans* retinal was added. The cells were incubated for 1 h at 25 °C, 200 rpm, before the expression of $pSRII^N$(KCWL)-Int^N-Strep was induced by adding 0.1 mM IPTG. The cells were incubated for 7 h at 25 °C, 200 rpm, with 10 μM all-*trans* retinal supplemented at 2 h-intervals. The cells were further incubated overnight (total ~20 h expression) at 18 °C, 200 rpm without supplementation of all-*trans* retinal, and harvested by centrifugation (5000 × *g*, 20 min, 4 °C).

Purification from the co-expression of $pSRII^N$(KCWL)-Int^N-Strep and Strep-Int^C-$pSRII^C$(KCWL)-His_6 was performed with the same method as for pSRII, except 5 mM TCEP-HCl was added to Buffer A, 2 mM TCEP-HCl to Buffer S, and 1 mM TCEP-HCl to all other buffers.

A.4.3 OmpX

Single transformed colonies from an LB agar plate were inoculated into 50 ml LB media with 50 μg/ml kanamycin and incubated overnight (37 °C, 180 rpm). Each overnight culture was then transferred to 500 ml M9 media [433] to give a starting OD_{600} = 0.1. Expression of OmpX as inclusion bodies was induced by adding 1 mM IPTG at an OD_{600} of 0.4–0.5. The cells were further incubated for 5 h at 37 °C, harvested by centrifugation (4000 × *g*, 4 °C, 20 min), and the cell pellets were frozen (–20 °C).

Inclusion body purification was done as previously described [377]. The harvested cell pellet was resuspended in Buffer TE and lysed by three passes through an Emulsiflex-C5 High Pressure Homogeniser (Avestin, Inc.) operating at 50 Psi. The solution was centrifuged (4300 × *g*, 4 °C, 1 h). Then, the pellet was homogenized (T18 basic ULTRA-TURRAX®, IKA®) in Buffer TE + 2% (v/v) Triton X-100 and shaken for 20 min at 37 °C. The solution

was centrifuged (4300 × *g*, 4 °C, 1 h), and the pellet was resuspended in Buffer TE, shaken for 1 h at 37 °C, and centrifuged again. The new pellet was dissolved in Buffer G and shaken for 2 h at 37 °C. Finally, the solution was centrifuged (14300 × *g*, 4 °C, 20 min), and the supernatant was retained.

The protein solution was concentrated using a 10 kDa MWCO centrifugal concentrator (Vivaspin) and loaded onto a size exclusion column equilibrated in Buffer U. Size exclusion chromatography was performed using either the Äkta Start (GE Healthcare) or the Äkta Pure (GE Healthcare) chromatography system. Either a Sephadex 200 10/300 GL column (GE Healthcare) running at 0.12 ml/min or a HiPrep 16/60 Sephacryl S-200 HR column (GE Healthcare) running at 0.5 ml/min was used. Typically, 1 ml and 4 ml fractions were collected from the respective columns. No significant differences were observed between the chromatography traces from the two columns. The relevant fractions were identified by SDS-PAGE, pooled, exchanged back into Buffer G using a PD-10 desalting column (GE Healthcare), and concentrated to about 300 μM.

Refolding was performed at room temperature by dropwise dilution of 5 ml 300 μM OmpX into 45 ml Buffer R with gentle stirring. The solution was stirred at room temperature overnight. The refolded protein was dialysed in a 3.5 kDa MWCO dialysis tubing (Spectrum Laboratories, Inc.) twice overnight against 5 L of Buffer D, followed by twice overnight against 5 L of Buffer N. 0.1% dodecylphosphocholine (DPC) was added to the protein solution between each change of buffer. Finally, the dialysed protein was concentrated using a 10 kDa MWCO centrifugal concentrator (Vivaspin).

A.4.4 OmpX in urea

^{15}N-OmpX in Buffer N with 10% D_2O and 5% DPC was first exchanged into Buffer U using a PD-10 desalting column. The eluted sample (4.5 ml) was loaded onto a HiPrep 16/60 Sephacryl S-200 HR column (GE Healthcare) running at 0.5 ml/min, and 3 ml fractions were collected. The fractions containing unfolded OmpX (assessed by SDS-PAGE) were pooled and dialysed twice for > 8 h each against 1 L of Buffer X. The dialysed sample was diluted/concentrated thrice in a 3 kDa MWCO centrifugal concentrator (Vivaspin) using 45 ml of Buffer X. The final DPC concentration, determined by 1D ^{1}H NMR, was < 0.02%.

A.5 Biophysics

A.5.1 UV/vis spectroscopy

UV/vis spectra were recorded at room temperature in a 10-mm path length quartz cuvette using a UV-1800 spectrophotometer (Shimadzu). Spectra were recorded between 250 and 600 nm with a fast scan rate (accumulation time = 0.05 s) and a sampling interval of 1 nm. Each steady-state spectrum was reported as an average of three measurements. Time-course measurements were done by recording spectra in 1-min or 5-min intervals using the aforementioned settings, with a typical dead-time of 25 s before the first spectrum was recorded.

Protein concentrations were confirmed by UV-vis spectroscopy using the method outlined above for steady-state spectra. Extinction coefficients (Table A.9) were obtained using ExPASy ProtParam tool (Swiss Institute of Bioinformatics) unless indicated otherwise. Concentrations of pSRII and OmpX were also confirmed by standard ion-exchange-ninhydrin analysis (performed by Peter Sharratt, Department of Biochemistry PNAC Facility, University of Cambridge).

Protein	Molecular weight (Da)	Wavelength (nm)	Extinction coefficient ($M^{-1}cm^{-1}$)
pSRII [434]	26420	498	48000
		280	49390
pSRII-KCWL	26967	280	54890
$pSRII^{N}$(KCWL)-Int^{N}-Strep	19870	280	39420
Strep-Int^{C}-$pSRII^{C}$(KCWL)-His_6	25747	280	42400
OmpX	16514	280	34840

Table A.9: Extinction coefficients of pSRII and OmpX

The purity of pSRII samples was calculated as follows:

$$\text{Purity (\%)} = \frac{A_{498}}{A_{280}} \times \frac{\varepsilon_{280}}{\varepsilon_{498}} \times 100 \qquad \text{(A.1)}$$

A.5.2 Far-UV CD

To avoid strong UV absorbance by NaN_3, NaN_3 was first removed from small aliquots (100 µl) of pSRII stock solutions by concentrating and diluting thrice in a 0.5-ml centrifugal filter unit (10-kDa cutoff) (Merck Millipore) using a total of ~1.5 ml of 50 mM sodium phosphate (pH 6.0), 50 mM NaCl and 0.1% c7-DHPC. CD spectra were recorded at 25 °C by scanning between 194 and 250 nm on an Aviv 410 spectrometer (Aviv Biomedical Inc.) using a 1×1-mm cuvette. Each sample contained 0.17 mg/ml protein in a volume of 400 µl, measured 2 h after exposure to unfolding buffer. CD spectra of the buffer solutions were subtracted from the sample spectra and smoothed using a window width of 11 and a degree of 2. Spectra were reported as the average of three scans. Data for each experimental condition are reported as the average of three independent samples.

All CD spectra were analysed using CDPro [298]. The helicity was estimated by taking basis set 10 containing both soluble and membrane proteins [298]. The built-in secondary structure-determination algorithms CONTIN/LL [298,435], SELCON3 [436] and CDSSTR [437] were executed, using delta epsilon as the input.

$$\Delta\varepsilon = \frac{\theta}{10 \times l \times c \times n \times 3298} \tag{A.2}$$

Here, θ is the observed ellipticity (millidegrees), l the path length (cm), c the protein concentration (M) and n the number of peptide bonds, taken as 246 for pSRII.

Molar residue ellipticity (MRE) is calculated as:

$$[\theta] = \Delta\varepsilon \times 3298 \tag{A.3}$$

To monitor time-dependent changes in α-helical content during SDS-mediated unfolding of pSRII, ellipticity was recorded at fixed wavelengths of 222 and 208 nm at 30-s intervals for 2 h. The typical experimental dead-time is 1 min.

A.5.3 Tryptophan Fluorescence

Fluorescence spectra were measured at 25 °C using a LS55 fluorescence spectrometer (Perkin Elmer) (excitation slit: 5.0 nm, emission slit: 17.5 nm). Emission scans were recorded

from 310 to 500 nm using an excitation wavelength of 295 nm, at a scan rate of 100 nm/min and with a sampling interval of 0.5 nm. For time-course measurements, fluorescence emission at 335 nm was recorded with an integration time of 3 s. The typical experimental dead-time was 25 s.

A.5.4 Unfolding pSRII in different denaturants

pSRII at ~450 μM was unfolded by a 67-fold dilution into unfolding buffer containing 50 mM sodium phosphate (pH 6.0), 50 mM NaCl, 0.1% c7-DHPC and different concentrations of denaturants. NaN_3 was excluded from unfolding buffers for fair data comparison across different biophysical techniques, as the strong UV absorbance by NaN_3 interferes with far-UV CD spectroscopy.

A.5.5 Hydroxylamine treatment of pSRII

Hydroxylamine hydrochloride was added to unfolding buffers to a final concentration of 10 mM. The relative concentrations of the monobasic and dibasic sodium phosphate salts were adjusted accordingly to ensure that the final pH values of all unfolding buffers containing hydroxylamine were consistent (within 0.1 pH unit) with pH values of matching conditions without hydroxylamine.

A.5.6 Refolding from SR_{440}

pSRII at ~450 μM was unfolded at room temperature for 4 min by 33-fold dilution into unfolding buffer containing 50 mM sodium phosphate (pH 6.0), 50 mM NaCl and 0.943 X_{SDS} (5% SDS + 0.5% c7-DHPC). The protein was refolded from the SR_{440} state by 12-fold dilution into refolding buffer containing 50 mM sodium phosphate (pH 6.0), 50 mM NaCl and different X_{SDS} in 0.5% c7-DHPC. The quoted X_{SDS} is the final X_{SDS} remaining in the reaction after 12-fold dilution of the unfolding reaction into refolding buffer. The refolding reaction was monitored by UV/vis spectroscopy for up to 2 h.

A.5.7 Refolding from SO_{390}

SO_{390} was obtained by a 2-h incubation of pSRII diluted 33-fold into unfolding buffer containing 50 mM sodium phosphate (pH 6.0), 50 mM NaCl and 0.893 X_{SDS} (2.5% SDS +

0.5% c7-DHPC). Refolding from SO_{390} was initiated in a similar manner as for refolding from SR_{440}, and the refolding reaction was monitored by UV/vis spectroscopy for up to 2 h. The final spectrum of the refolded sample was recorded at least 16 h after refolding was initiated.

A.5.8 Calculations

A.5.8.1 Mole fraction of SDS

The mole fraction of SDS (X_{SDS}) was calculated as the bulk detergent composition in solution.

$$X_{SDS} = \frac{[\text{SDS}]}{[\text{SDS}]+[\text{c7-DHPC}]} \quad \text{(A.4)}$$

A.5.8.2 Equilibrium unfolding

Absorbance at 498 nm was plotted against X_{SDS}. Assuming that only the folded and unfolded conformations are present at any given time point of the unfolding reaction, the fraction of protein present in the folded (f_F) and unfolded (f_U) conformations, respectively, can then be represented as:

$$f_F + f_U = 1 \quad \text{(A.5)}$$

f_U was calculated as:

$$f_U = (A_F - A_i)/(A_F - A_U) \quad \text{(A.6)}$$

where A_F is the A_{498} of completely folded or native protein, A_i is the observed A_{498} at any point of denaturant concentration, A_U is A_{498} of the completely denatured protein.

A.5.9 Kinetics fitting

UV/vis spectroscopy, CD spectroscopy and tryptophan fluorescence time-course data were subjected to Levenberg-Marquardt nonlinear least squares fitting using R codes written by James Mitchell (University of Warwick) [438].

A.5.9.1 UV/Vis spectra

Time courses of absorbance spectra were modelled as the combinations of absorbance peaks varying over time according to a kinetic model. Each peak was treated as a Gaussian peak of the form $e^{-\frac{(\lambda-c)^2}{2w^2}}$, where c is the wavelength corresponding to the fitted peak centre and w is the half width at half maximum of the fitted Gaussian.

All absorbances in the region 350–650 nm were assumed to be due to retinal species varying according to the kinetic model. The kinetic model was described with two consecutive first-order reactions, converting pSRII (absorbing at 498 nm) to SR_{440}, which subsequently decays to SO_{390}:

$$\mathrm{pSRII} \xrightarrow{k_\mathrm{u}} \mathrm{SR}_{440} \xrightarrow{k_\mathrm{h}} \mathrm{SO}_{390} \tag{A.7}$$

Time courses were unfolded to a three-column matrix with a column for wavelength (nm), time (min) and absorbance for that wavelength at that time. These were then fitted to Equation A.8 using the Levenberg-Marquardt least-squares fitting algorithm [250] as implemented in the minpack.LM package [439] for R [440].

$$A \approx \varepsilon_{498}\left([Total]e^{-k_\mathrm{u}(t+f)}\right)e^{-\frac{(\lambda-c_{498})^2}{2w_{498}{}^2}} + \varepsilon_{440}[Total]\left(\frac{k_\mathrm{u}}{k_\mathrm{h}-k_\mathrm{u}}\right)\left(e^{-k_\mathrm{u}(t+f)} - e^{-k_\mathrm{h}(t+f)}\right)e^{-\frac{(\lambda-c_{440})^2}{2w_{440}{}^2}} + \varepsilon_{360}[Total]\left(1+\frac{1}{k_\mathrm{u}-k_\mathrm{h}}\left(k_\mathrm{h}e^{-k_\mathrm{u}(t+f)} - k_\mathrm{u}e^{-k_\mathrm{h}(t+f)}\right)\right)e^{-\frac{(\lambda-c_{390})^2}{2w_{390}{}^2}} \tag{A.8}$$

where [Total] is the total protein concentration (see Equation A.9), k_u and k_h are the fitted rate constants, and f is an error term that is added to allow for the dead-time and the small amount of absorbance at 380 nm present in native spectra. Rate constants were reported as the mean of three independent measurements, with error bars representing the standard deviation.

As no retinal was added during the experiment, at every point, the Beer-Lambert law was used with a known total concentration to approximate extinction coefficient-like values (Equation A.9) to convert the amplitudes of the fitted peaks to concentrations while accounting for peak overlap. This allowed peak amplitudes to be replaced in fitting with the concentration terms of the rate equations.

$$[\mathrm{Total}] \approx \frac{I_{498}}{\varepsilon_{498}} + \frac{I_{440}}{\varepsilon_{440}} + \frac{I_{390}}{\varepsilon_{390}} \tag{A.9}$$

A.5.9.2 Tryptophan fluorescence and CD

Rate equations were used to fit tryptophan fluorescence (intensity at 335 nm vs. time) and CD (MRE at 222 nm vs. time) time-courses. Fluorescence increases over time were fit to the sum of the rate equations for SR_{440} and SO_{390}. The initial fluorescence was subtracted to

remove contribution from pSRII and simplify fitting. This yields Equation A.10 when k_1 is sufficiently large.

$$\Delta\text{Intensity} = I_{\text{Prot}}[\text{Total}]\left(\frac{k_{\text{u}}}{k_{\text{h}}-k_{\text{u}}}\right)(e^{-k_{\text{u}}t} - e^{-k_{\text{h}}t}) + I_{\text{Apo}}[\text{Total}]\left(1 + \left(\frac{1}{k_{\text{u}}-k_{\text{h}}}\right)(k_{\text{h}}e^{-k_{\text{u}}t} - k_{\text{u}}e^{-k_{\text{h}}t})\right) \quad \text{(A.10)}$$

Under conditions where $k_{\text{u}} < k_{\text{h}}$, time-courses approximate a single exponential curve. Single and double exponential fits to the data were compared using the Akaike Information Criterion [268] and checked with cross-validation wherever ambiguous.

A.5.9.3 Refolding from SR_{440}

$$\text{A}(\lambda) \sim \text{D} + \frac{E\times 10^7}{\lambda^G} + \text{A}_{390} \times e^{-\frac{(\lambda-c_{390})^2}{2w_{390}{}^2}} + \text{A}_{440} \times e^{-\frac{(\lambda-c_{440})^2}{2w_{440}{}^2}} + \text{A}_{500} \times e^{-\frac{(\lambda-c_{500})^2}{2w_{500}{}^2}} \quad \text{(A.11)}$$

The 350- to 650-nm region of the absorption spectra of pSRII was fitted to Equation A.11. Each time point of the refolding time-courses was fitted to a combination of three Gaussian peaks at around 390, 440 and 500 nm, and terms for both linear and scattering baseline variation. Changes in the fitted absorbance intensities at ~500 nm over time were fitted to a double exponential model:

$$A_{500} \sim A_1(1 - e^{-R_1 t}) + A_2(e^{-R_2 t}) + c \quad \text{(A.12)}$$

where R_1 and R_2 are the macroscopic rate constants, and A_1 and A_2 are the amplitudes of the fast phase and the slow phase, respectively.

R_1 and R_2 are the solutions to the following quadratic equation:

$$R^2 - (k_f + k_u + k_h)R + k_u k_h = 0 \quad \text{(A.13)}$$

Derivation of Equation A.13 was done as outlined by Schlebach *et al.* [118].

Based on the quadratic equation, the microscopic rate constants k_{f} and k_{u} were extracted using the following relationships and experimentally determined values of R_1, R_2 and k_{h}:

$$R_1 R_2 = k_u k_h \quad \text{(A.14)}$$

$$R_1 + R_2 = k_f + k_u + k_h \quad \text{(A.15)}$$

Under conditions where the rate of retinal hydrolysis is very slow compared to the rate of folding, changes in the fitted absorbance intensities at ~500 nm over time were approximated to a single exponential model:

$$A_{500} \sim A_1\left(1 - e^{-k_f t}\right) + c \quad \text{(A.16)}$$

where the observed rate constant was approximated as the *de novo* refolding rate k_f.

A.5.9.4 Refolding from SO_{390}

The absorption spectrum of native pSRII contains one main retinal absorption peak at 498 nm, three shoulder peaks at 460, 420 and 370 nm, and protein absorption peak at 280 nm [224]. Five native spectra were fitted to these peaks, four of which were symmetrical Gaussian functions, and the 420-nm peak, the width of which was allowed to vary sigmoidally with wavelength to allow for previously reported skewedness [224]. Retinal peak amplitudes were normalised to have a sum of 1, and parameters were averaged for the retinal peaks to find a peak profile for native pSRII.

Each time point of the refolding time-courses was fitted as a combination of some multiple of the native profile, a peak around 390 nm for free retinal and terms for both linear and scattering baseline variation (see the first two terms of Equation A.17):

$$\mathrm{A}(\lambda) \sim \mathrm{D} + \frac{E \times 10^7}{\lambda^G} + \mathrm{A}_{280} \times e^{-\frac{(\lambda - c_{280})^2}{2w_{280}{}^2}} + \mathrm{A}_{390} \times e^{-\frac{(\lambda - c_{390})^2}{2w_{390}{}^2}} + \mathrm{B} \times (6.600146 \times 10^{-2} e^{-\frac{(\lambda - 371.1319)^2}{(2 \times 15.86762)^2}} + 2.151851 \times 10^{-1} e^{-\frac{(\lambda - 438.7399)^2}{(2 \times 32.81915/(1 + e^{(9.347848 \times 10^{-3} \times (\lambda - 438.7399))}))^2}} + 1.368525 \times 10^{-1} e^{-\frac{(\lambda - 461.5911)^2}{(2 \times 7.163922)^2}} + 5.819610 \times 10^{-1} e^{-\frac{(\lambda - 496.4147)^2}{(2 \times 16.53747)^2}}) \quad \text{(A.17)}$$

The concentrations of free retinal and native pSRII were extracted from these time-points. Since the total concentration is constant, and the parameters B and A_{390} were assumed to be proportional to the concentrations of native pSRII and free retinal respectively, the concentrations of these species were estimated by fitting these parameters for each time-point to Equation A.18.

$$[\mathrm{Total}] = \mathrm{A}_{390} \times \mathrm{SO}_{390} + \mathrm{B} \times \mathrm{pSRII} \quad \text{(A.18)}$$

where [Total] is the total protein concentration determined from the A_{280} peak, A_{390} and B were fitted values from Equation A.18, and SO_{390} and pSRII were fitted coefficients, later used to calculate concentrations.

The estimated values for SO_{390} concentration were fitted to an exponential decay model of the form $A \times e^{-k_{\mathrm{app}} t} + C$, from which an apparent rate constant, k_{app}, and an asymptote, C,

were obtained. The asymptote of the curve estimates the fraction of the total concentration that cannot be refolded on the timescale of the experiment (2 h).

A.6 NMR spectroscopy

All pulse sequences were written by Dr. Daniel Nietlispach. A final concentration of 10% D_2O (v/v) was typically (unless stated otherwise) added to the protein sample before NMR experiment to provide lock stabilisation. All ^1H chemical shifts were referenced to TSP (trimethylsilylpropanoic acid) as internal standard. Detergent concentrations of all protein samples were confirmed by comparing the detergent peaks in the 1D ^1H NMR spectra of the protein sample vs. a solution of known detergent concentration.

A.6.1 1D ^1H NMR

1D ^1H NMR spectra were measured using a Bruker AvanceIII AV800 equipped with 5 mm TXI CryoProbe (HCN/z). Pulse sequences were 1D spin echos. Water suppression was achieved using a 3-9-19 180° WATERGATE sequence. All spectra of pSRII were recorded at 308 K, for 512 complex data points, 512 scans and acquisition time of 51.2 ms, with a spectral width of 10000 Hz. Selective suppression of detergent signals was achieved using a Gaussian Q5 cascade with duration of 5 ms followed by purging by gradient pulses. Spectra of 0.2% DPC with 5% D_2O and OmpX in Buffer X with 5% D_2O were recorded at 293 K and 288 K, respectively. Both spectra were recorded for 9999 complex data points, 4 scans and acquisition time of 1 s, and with spectral width of 10000 Hz.

A.6.2 1D ^{19}F NMR

1D ^{19}F NMR spectra were measured using a Bruker AvanceIII AV600 spectrometer equipped with a 5 mm QCI CryoProbe (HCNF/z). All ^{19}F NMR chemical shifts were referenced to $CFCl_3$ (trichloro-fluoro-methane) as external standard. The pulse sequence was a spin-echo. All spectra were recorded with ^1H broadband decoupling with 4997 complex data points, 1024 scans and acquisition time of 100 ms.

A.6.3 Multidimensional (2D and 3D) NMR

A.6.3.1 pSRII

All 2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC experiments were recorded with samples of ^{15}N-labelled pSRII in 50 mM NaH_2PO_4/Na_2HPO_4, 50 mM NaCl, 0.5% c7-DHPC with varied pH and SDS concentrations, at 308 K. Except for the two experiments on ^{15}N,^{19}F-5-Trp-apo-pSRII, all spectra were recorded on a Bruker AvanceIII AV800 spectrometer equipped with a 5 mm TXI CryoProbe (HCN/z), with 512 × 128 complex points, spectral width of 2777.78 Hz in the indirect dimension (t_{1max} = 46.08 ms) and acquisition time of 51.2 ms. For ^{15}N,^{19}F-5-Trp-apo-pSRII, experiments were using a Bruker AvanceIII AV600 spectrometer equipped with a 5 mm QCI CryoProbe (HCNF/z), with 512 × 100 complex points, spectral width of 2500 Hz in the indirect dimension (t_{1max} = 52.70 ms) and acquisition time of 51.2 ms. Frequency discrimination in the ^{15}N dimension was achieved using P/N-type coherence order selection combined with axial peak shifting by one-half the spectral width. Solvent suppression was not required due to utilisation of amide-selective ^{1}H pulses for magnetisation transfer.

Details of the sample concentration and number of scans for each reported spectrum are shown in Table A.10.

^{15}N R_1 and $R_{1\rho}$ experiments were recorded at 308 K on a Bruker AvanceIII AV600 spectrometer equipped with a 5 mm QCI CryoProbe (HCNF/z) using samples of ~525 μM ^{15}N-pSRII in 0 X_{SDS} (2% c7-DHPC) and in 0.728 X_{SDS} (0.8% SDS + 0.5% c7-DHPC). Sample preparation and NMR experiments on ^{15}N-pSRII in 0 X_{SDS} (2% c7-DHPC) were performed by Dr. Mark Bostock. TROSY-modified pulse sequences were used. All spectra were recorded with 1024 × 64 complex points. Both ^{15}N $R_{1\rho}$ experiments were recorded using an adiabatically ramped spin lock field with a B_1 field strength of 1164 Hz. On- and off-resonance versions of the experiment were recorded (Table A.11).

Sample	X_{SDS} (% SDS)	[c7-DHPC] (%)	D_2O (%)	pH	[pSRII] (µM)	No. of Scans
^{15}N,^{19}F-5-Trp-pSRII	0	1.8	10	6.0	157	80
^{15}N,^{19}F-5-Trp-apo-pSRII	0	1.8			162	96
	0.625 (1.0%)	1			90	56
^{15}N-pSRII	0	0.5			83	160
	0.400 (0.2%)				94	240
	0.625 (0.5%)				94	
	0.728 (0.8%)				94	
	0.769 (1.0%)				94	
	0.769 (2.0%)	0.9			94	
	0.834 (1.5%)	0.5			83	
	0.870 (2.0%)				94	
	0.893 (2.5%)				94	
	0.909 (3.0%)				83	
	0.971 (10%)				83	
	0.985 (20%)				83	
	0.990 (30%)				83	
	0		10		250	
			90			
	0.769 (1.0%)		10		240	
			90			
	0		10	4.7	112	
				3.7		
				2.6		
				2.0		

Table A.10: Data acquisition parameters for NMR experiments on ^{15}N-pSRII in different X_{SDS} and at different pHs

Sample	Experiment	No. of delays	t_{max} (ms)	No. of Scans	Offset (Hz)
0 X_{SDS} (2% c7-DHPC)	^{15}N R_1	9	3040	32	-
	^{15}N $R_{1\rho}$	10	200	128	1500
0.728 X_{SDS} (0.8% SDS + 0.5% c7-DHPC)	^{15}N R_1	6	2000	120	-
	^{15}N $R_{1\rho}$	8	100	128	0

Table A.11: Data acquisition parameters for ^{15}N relaxation experiments.
t_{max} represents the maximum relaxation delay.

A.6.3.2 RalAΔC-GDP

The sample of 700 μM ^{15}N-labelled RalAΔC-GDP in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM $MgCl_2$, 0.05% NaN_3 and 10% D_2O was made by Dr. Arooj Shafiq. Descriptions of the pulse sequences for IPAP-HSQC and $^1J_{NH}$-modulated HSQC (and variations thereof) can be found in Chapter 4. All experiments were recorded at 298 K on a Bruker Avance DRX500 spectrometer equipped with 5 mm TCI CryoProbe (HCN/z).

Experiment	Complex points (^{1}H, ^{15}N)	t_{1max} (ms)	Acquisition time (ms)	No. of scans	$^1J_{NH}$ evolution (ms)
IPAP-HSQC	512 × 300	136.2	51.2	32	-
				112	-
SOFAST-IPAP-HSQC	512 × 300	136.2	51.2	64	-
$^1J_{NH}$-modulated HSQC	512 × 128	58.11	51.2	8	11.60–37.20
				16*	11.60–37.20
				32	11.60–37.20
				32	32.88–58.48
				8*	32.88–58.48
$^1J_{NH}$-modulated TROSY-HSQC	1024 × 128	58.11	102.4	64*	32.88–58.48
				48*	22.24–47.84

Table A.12: Data acquisition parameters for NMR experiments on ^{15}N-RalAΔC-GDP

A.6.3.3 OmpX

All experiments were recorded on a Bruker AvanceIII AV800 spectrometer equipped with 5 mm TXI CryoProbe (HCN/z).

995 µM ^{15}N-labelled OmpX 20 mM NaPi (pH 6.8), 100 mM NaCl, 5 mM EDTA, 3.8% DPC and 10% D_2O Temperature: 318 K					
Experiment	**Complex points (^{1}H, ^{15}N, ^{1}H)**	**t_{1max} (ms)**	**Acquisition time (ms)**	**No. of Scans**	**$^1J_{NH}$ evolution (ms)**
3D ^{15}N-separated NOESY-HSQC	512 × 36 × 110	t_{1max}: 12.00 ms t_{2max}: 11.95 ms	51.2	24	-
2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC	512 × 150	45.00	51.2	32	-
$^1J_{NH}$-modulated HSQC	512 × 100	40.84	51.2	40	11.60–37.20
	512 × 100	40.84	51.2	96	11.60–37.20
$^1J_{NH}$-modulated TROSY-HSQC	512 × 100	40.84	51.2	40	11.60–37.20
SOFAST-IPAP-HSQC	512 × 150	61.26	51.2	64	-
221 µM ^{15}N-labelled OmpX 20 mM NaPi (pH 6.5), 8 M urea, 0.1 mM NaN_3, 5% D_2O Temperature: 288 K					
2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC	512 × 128	46.08	51.2	80	-
SOFAST-IPAP-HSQC	1024 × 300	150.0	102.4	64	-
$^1J_{NH}$-modulated HSQC	1024 × 200	100.0	102.4	40	22.24–47.84
166 µM ^{15}N-labelled OmpX 20 mM NaPi (pH 6.5), 6 M urea, 0.1 mM NaN_3, 5% D_2O Temperature: 288 K					
2D [^{1}H,^{15}N] SOFAST-TROSY-HSQC	512 × 128	46.08	51.2	80	-
SOFAST-IPAP-HSQC	1024 × 300	150.0	102.4	64	-

Table A.13: Data acquisition parameters for NMR experiments on ^{15}N-OmpX

A.6.4 Data Analysis

Raw NMR data were processed in Azara v2.8 (Dr. Wayne Boucher, Department of Biochemistry, University of Cambridge). Assignment and measurement of peak intensities were performed in CCPN Analysis [375].

A.6.4.1 Chemical shift differences

Chemical shift differences were calculated using Equation A.19:

$$\Delta\delta(^{1}\mathrm{H}, ^{15}\mathrm{N}) = \sqrt{\Delta\delta(^{1}\mathrm{H})^{2} + \frac{\Delta\delta(^{15}\mathrm{N})^{2}}{6}} \tag{A.19}$$

A.6.4.2 $^{1}J_{\mathrm{NH}}$ coupling constants

Using custom-written R code (from Dr. Mark Bostock), $^{1}J_{\mathrm{NH}}$ scalar coupling constants (denoted as J) were obtained from $^{1}J_{\mathrm{NH}}$-modulated HSQC experiments using a fit to Equation A.20:

$$I(\tau) = A\sin(\pi J\tau)\, e^{-C\tau} \tag{A.20}$$

where $I(\tau)$ denotes the peak intensity, τ the variable delay time in seconds, C the decay rate of the ^{15}N magnetisation as a result of transverse relaxation, and A is an arbitrary signal intensity scaling factor.

Signal-to-noise ratios (SNR) were calculated as:

$$\mathrm{SNR} = \frac{I}{\sigma_n} \tag{A.21}$$

where I denotes the peak intensity, and σ_n the standard deviation of the noise.

Root mean squared deviation in peak position ($\mathrm{RMSD_{pos}}$) was calculated as:

$$\mathrm{RMSD_{pos}} = \frac{1}{\mathrm{SNR}}\left(\frac{0.26}{t_{1\max}} + 0.067R_2\right) \tag{A.22}$$

where SNR denotes the signal-to-noise ratio of a peak calculated using Equation A.21, $t_{1\max}$ the maximum increment of the t_1 evolution period in seconds, and R_2 the transverse relaxation rate constant. R_2 for individual residues was taken as C in Equation A.20, obtained by fitting peak intensities from $^{1}J_{\mathrm{NH}}$-modulated HSQC experiments. Note that these R_2 values are free-evolution R_2* values.

For IPAP-HSQC experiments, the RMSD in $^{1}J_{\mathrm{NH}}$ of each residue was calculated as the sum of $\mathrm{RMSD_{pos}}$ for each of the two peaks in the ^{15}N doublet.

Appendix

B Chemical shifts and intensity ratios for pSRII

B.1 Backbone chemical shifts and intensity ratios for pSRII in SDS

Table B.1: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0–0.728 X_{SDS} (0–0.8% SDS)
Conditions: ~90 µM ^{15}N-pSRII, 50 mM sodium phosphate (pH 6.0), 50 mM NaCl, 0.5% c7-DHPC, 10% D_2O, 308 K.
Alternative conformations are labelled with letters.

Table B.1: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0–0.728 X_{SDS} (0–0.8% SDS + 0.5% c7-DHPC)

#	Resi	0 X_{SDS}		0.400 X_{SDS} (0.2% SDS)			0.625 X_{SDS} (0.5% SDS)			0.728 X_{SDS} (0.8% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
2	Val	7.948	120.186	7.877	119.583	0.910	7.824	119.029	0.696	7.776	118.686	0.654
3	Gly	8.507	112.366	8.461	112.335	0.927	8.379	112.269	0.934	8.321	112.221	0.923
5	Thr	8.339	108.141	8.266	107.931	0.835	8.218	107.754	0.634	8.189	107.670	0.672
6	Thr	7.479	117.585	7.408	117.463	0.817	7.365	117.354	0.944	7.347	117.250	0.650
7	Leu	7.350	120.064	7.316	119.825	0.933	7.289	119.685	0.777	7.263	119.577	0.868
9	Trp	8.641	119.955	8.616	119.874	0.749	8.597	119.845	0.687	8.588	119.812	0.663
10	Leu	8.209	120.260									
11	Gly	7.970	105.314	7.965	105.249	1.088	7.944	105.178	0.970	7.936	105.111	0.988
12	Ala	8.378	123.335	8.351	123.302	1.196	8.346	123.271	0.722	8.347	123.282	0.406
13	Ile	8.876	116.576	8.859	116.558	0.686	8.848	116.545	0.700	8.843	116.554	0.534
14	Gly	8.702	107.748	8.699	107.692	0.777	8.699	107.616	0.750	8.696	107.573	0.617
15	Met	8.352	117.820	8.335	117.751	0.673	8.319	117.671	0.745	8.308	117.607	0.582
16	Leu	8.843	125.236	8.828	125.291	0.752	8.815	125.305	0.748	8.812	125.330	0.638

Table B.1: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0–0.728 X_{SDS} (0–0.8% SDS + 0.5% c7-DHPC)

#	Resi	0 X_{SDS}		0.400 X_{SDS} (0.2% SDS)			0.625 X_{SDS} (0.5% SDS)			0.728 X_{SDS} (0.8% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
17	Val	8.506	119.005	8.520	119.056	0.713	8.526	119.035	0.681	8.529	119.015	0.591
17b	Val									8.567	119.138	0.157
18	Gly	7.813	104.255	7.800	104.133	0.732	7.792	104.097	0.714	7.785	104.053	0.590
20	Leu	8.089	118.821	8.108	118.747	0.379	8.104	118.675	0.589			
21	Ala	8.015	121.350	8.027	121.356	0.851	8.045	121.360	0.639	8.048	121.340	0.583
22	Phe	8.591	118.274	8.627	118.324	0.800	8.642	118.217	0.835	8.645	118.187	0.665
23	Ala	8.420	121.316	8.454	121.394	0.868	8.464	121.506	0.654	8.471	121.565	0.584
24	Trp	8.651	118.459	8.685	118.455	0.828	8.707	118.461	0.712	8.726	118.495	0.645
25	Ala	8.628	120.971	8.656	120.996	0.788	8.719	121.087	0.715	8.750	121.160	0.605
26	Gly	7.642	103.903	7.714	104.182	0.841	7.749	104.336	0.718	7.761	104.405	0.657
27	Arg	7.176	120.777	7.182	120.796	0.794	7.180	120.813	0.763	7.184	120.845	0.698
28	Asp	7.804	118.024	7.733	117.784	0.734	7.712	117.779	0.705	7.709	117.810	0.641
29	Ala	7.157	123.269	7.160	123.139	0.809	7.161	123.058	0.850	7.168	123.059	0.795
30	Gly	8.765	109.747	8.759	109.757	0.831	8.745	109.804	0.861	8.728	109.830	0.793
31	Ser	8.664	115.546	8.645	115.770	1.300	8.624	115.887	1.520	8.603	115.912	1.471
32	Gly	8.998	111.170	8.912	111.657	1.369	8.843	112.324	1.623	8.811	112.771	1.527
35	Arg	8.419	115.863	8.477	115.986	1.129	8.489	116.146	1.101	8.497	116.340	0.514
37	Tyr	7.193	116.858	7.224	116.696	0.742	7.233	116.665	0.568	7.223	116.593	0.473
38	Val	8.643	117.903	8.601	117.727	0.719	8.515	117.459	0.697	8.497	117.339	0.495
40	Leu	7.310	118.971	7.206	119.004	1.079	7.149	119.034	1.031	7.122	119.022	0.946
41	Val	8.153	118.234	8.152	118.345	0.383						
42	Gly	8.417	106.641	8.406	106.528	0.709	8.368	106.408	0.714	8.355	106.341	0.545
43	Ile	8.116	120.539	8.102	120.497	0.578						
44	Ser	7.844	109.747	7.867	109.800	0.701	7.888	109.844	0.687	7.890	109.819	0.630
45	Gly	8.898	114.911	8.880	114.897	0.808	8.858	114.849	0.743	8.853	114.868	0.668

Table B.1: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0–0.728 X_{SDS} (0–0.8% SDS + 0.5% c7-DHPC)

#	Resi	0 X_{SDS}		0.400 X_{SDS} (0.2% SDS)			0.625 X_{SDS} (0.5% SDS)			0.728 X_{SDS} (0.8% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
46	Ile	8.062	120.636	8.072	120.649	0.811	8.072	120.631	0.931	8.072	120.625	1.012
47	Ala	6.689	120.852	6.693	120.748	0.805	6.701	120.710	0.811	6.692	120.696	0.629
48	Ala	8.484	120.132	8.474	120.249	0.762	8.467	120.298	0.702	8.462	120.345	0.619
49	Val	7.503	115.657	7.512	115.726	0.905	7.522	115.776	0.768	7.522	115.815	0.578
50	Ala	8.397	122.457	8.404	122.458	0.812	8.411	122.481	0.792	8.413	122.482	0.632
51	Tyr	8.801	115.439	8.798	115.511	0.741	8.792	115.545	0.656	8.792	115.571	0.537
52	Ala	8.132	121.707	8.155	121.887	0.735	8.175	122.007	0.688	8.188	122.085	0.606
53	Val	8.268	113.936	8.300	114.146	0.753	8.325	114.255	0.735	8.339	114.341	0.574
54	Met	7.498	118.559	7.509	118.382	1.043	7.517	118.357	0.896	7.534	118.381	1.261
55	Ala	8.330	124.951	8.329	124.930	2.178	8.358	125.083	1.448	8.376	125.199	1.127
56	Leu	7.262	115.991	7.286	116.041	0.695	7.324	116.104	0.842	7.335	116.144	1.036
57	Gly	8.035	107.243	8.040	107.241	0.786	8.050	107.372	0.795	8.052	107.466	0.656
58	Val	7.620	122.869	7.629	122.903	0.776	7.646	123.027	0.734	7.661	123.149	0.616
59	Gly	8.834	112.230	8.574	111.893	0.672	8.429	111.664	0.713	8.362	111.529	0.639
60	Trp	8.464	122.785	8.461	122.829	0.768	8.440	122.802	0.675	8.420	122.756	0.598
61	Val	9.028	127.097	9.005	127.312	0.896	8.993	127.415	0.894	8.989	127.460	0.779
63	Val	8.435	126.326	8.402	126.389	0.899	8.374	126.411	0.954	8.357	126.424	0.751
64	Ala	8.835	128.446	8.848	128.364	0.847	8.845	128.295	0.722	8.836	128.247	0.610
65	Glu	8.446	120.820									
66	Arg	8.605	121.424	8.635	121.294	0.893	8.653	121.201	0.912	8.668	121.174	0.748
67	Thr	8.010	118.069	8.019	118.086	0.704	8.021	118.132	0.851	8.022	118.199	0.950
68	Val	9.210	128.868	9.234	129.054	0.855	9.254	129.193	0.865	9.264	129.280	0.725
69	Phe	8.018	127.209	8.021	127.302	0.907	8.024	127.366	1.571	8.024	127.410	1.234
70	Val	8.575	123.674	8.554	123.461	0.959	8.537	123.350	0.917	8.520	123.298	0.807
74	Ile	7.025	117.767	6.992	117.714	0.839	6.976	117.682	0.881	6.965	117.668	0.688

Table B.1: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0–0.728 X_{SDS} (0–0.8% SDS + 0.5% c7-DHPC)

#	Resi	0 X_{SDS}		0.400 X_{SDS} (0.2% SDS)			0.625 X_{SDS} (0.5% SDS)			0.728 X_{SDS} (0.8% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
75	Asp	7.607	117.374	7.593	117.485	0.773	7.584	117.507	0.760	7.570	117.509	0.677
76	Trp	7.940	115.164	7.887	115.024	1.237	7.843	114.885	2.332	7.822	114.777	4.547
77	Ile	8.280	120.176	8.279	120.149	0.806						
78	Leu	7.504	115.756	7.500	115.767	0.840	7.483	115.739	0.538	7.474	115.705	0.425
79	Thr	9.791	107.411	9.788	107.280	0.741	9.782	107.188	0.816	9.781	107.133	0.638
80	Thr	8.186	117.719	8.192	117.736	0.845	8.191	117.744	0.661	8.197	117.760	0.377
82	Leu	6.788	117.856	6.762	117.814	0.731	6.753	117.742	0.747	6.756	117.727	0.651
83	Ile	7.778	122.669	7.811	122.815	0.934	7.831	122.867	0.777	7.843	122.919	0.604
85	Tyr	8.990	122.438	9.011	122.559	0.670	9.029	122.689	0.745	9.038	122.762	0.573
86	Phe	8.550	121.521	8.590	121.628	0.587	8.616	121.605	0.393	8.630	121.640	0.448
87	Leu	8.466	116.583	8.511	116.762	0.761	8.537	116.887	0.586	8.547	116.945	0.535
88	Gly	9.360	108.206	9.387	108.287	0.737	9.413	108.357	0.744	9.424	108.397	0.532
89	Leu	8.892	126.017	8.971	126.434	0.911	9.024	126.709	0.837	9.048	126.826	0.702
90	Leu	7.666	119.836	7.714	119.955	0.703	7.743	120.123	0.768	7.750	120.123	0.756
91	Ala	8.100	116.884	8.140	117.078	0.523	8.167	117.162	0.633	8.176	117.216	0.516
92	Gly	7.562	105.638	7.588	105.941	0.810	7.606	106.156	0.861	7.612	106.235	0.749
93	Leu	7.253	118.078	7.262	118.099	0.418	7.271	118.040	0.398	7.265	118.014	0.388
94	Asp	8.404	120.824	8.359	120.754	0.852	8.311	120.743	0.717	8.289	120.721	0.631
95	Ser	8.572	114.953	8.483	114.501	0.738	8.423	114.286	0.728	8.395	114.187	0.677
96	Arg	8.231	123.224	8.092	122.869	0.852	8.000	122.716	0.829	7.957	122.585	0.648
97	Glu	8.244	119.403	8.264	119.400	0.537						
98	Phe	8.531	118.132	8.552	117.891	0.887	8.559	117.731	0.729	8.560	117.651	0.638
99	Gly	8.187	105.159	8.251	105.376	0.794	8.302	105.436	0.768	8.321	105.435	0.596
100	Ile	7.779	123.485	7.884	123.436	0.724	7.961	123.398	0.700	7.989	123.375	0.549
101	Val	7.675	115.946	7.717	115.956	0.825	7.738	116.147	0.718	7.736	116.225	0.587

Table B.1: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0–0.728 X_{SDS} (0–0.8% SDS + 0.5% c7-DHPC)

#	Resi	0 X_{SDS}		0.400 X_{SDS} (0.2% SDS)			0.625 X_{SDS} (0.5% SDS)			0.728 X_{SDS} (0.8% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
102	Ile	8.538	120.582	8.592	120.880	0.708	8.626	121.124	0.885	8.633	121.220	0.711
103	Thr	7.961	122.658	7.957	122.690	0.767	7.941	122.687	0.734	7.937	122.668	1.000
105	Asn	8.047	115.250	8.036	115.201	0.770	8.018	115.137	0.802	8.006	115.086	0.641
106	Thr	8.072	114.555	8.086	114.651	0.803	8.102	114.743	0.643	8.109	114.794	0.515
108	Val	7.581	118.040	7.552	117.948	0.692	7.527	117.879	0.756	7.514	117.842	0.599
109	Met	7.395	112.681	7.430	112.726	0.696	7.437	112.840	0.659	7.433	112.927	0.585
110	Leu	8.839	118.819	8.866	119.043	0.897	8.887	119.202	0.889	8.895	119.286	0.705
111	Ala	8.529	121.369	8.514	121.378	0.734	8.505	121.387	0.759	8.501	121.393	0.647
112	Gly	8.121	102.804	8.154	102.848	0.854	8.171	102.850	0.899	8.176	102.855	0.613
113	Phe	7.982	122.037	7.986	122.083	0.678	8.006	122.101	2.051	7.974	122.088	2.179
114	Ala	8.724	119.281	8.732	119.393	0.776	8.742	119.490	0.772	8.746	119.564	0.613
115	Gly	8.542	101.318	8.562	101.284	0.807	8.576	101.261	0.801	8.584	101.278	0.639
116	Ala	7.808	121.466	7.806	121.478	0.800	7.805	121.462	0.856	7.804	121.469	0.819
117	Met	7.120	112.781	7.134	112.946	0.725	7.150	113.130	0.699	7.164	113.270	0.540
118	Val	7.208	121.328	7.216	121.448	0.727	7.216	121.427	0.643	7.216	121.379	0.547
121	Ile	8.590	120.363	8.577	120.665	0.863	8.568	120.810	0.885	8.556	120.882	0.739
122	Glu	10.303	122.575	10.184	122.105	0.817	9.951	121.534	0.886	9.792	121.137	0.571
123	Arg	8.102	120.558	8.097	120.512	0.834						
124	Tyr	8.025	118.968	7.997	118.783	0.928	8.055	118.811	1.079	8.081	118.784	1.006
125	Ala	7.836	122.290	7.844	122.127	0.952	7.834	122.051	0.963	7.817	122.008	0.748
126	Leu	6.989	118.440	6.941	118.350	0.592	6.905	118.279	0.561	6.898	118.263	0.491
127	Phe	7.899	118.978									
128	Gly	8.410	105.675	8.395	105.587	0.809	8.379	105.557	0.908	8.369	105.559	0.825
129	Met	7.867	120.523	7.873	120.771	0.599	7.872	120.939	0.615	7.867	121.019	0.456
131	Ala	8.557	124.210	8.546	124.237	0.855	8.535	124.262	0.807	8.527	124.271	0.634

Table B.1: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0–0.728 X_{SDS} (0–0.8% SDS + 0.5% c7-DHPC)

#	Resi	0 X_{SDS}		0.400 X_{SDS} (0.2% SDS)			0.625 X_{SDS} (0.5% SDS)			0.728 X_{SDS} (0.8% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
132	Val	7.911	116.923	7.915	116.866	0.731	7.924	116.872	0.517	7.927	116.863	0.501
133	Ala	7.784	121.009	7.779	121.045	0.962	7.795	121.058	0.861	7.799	121.095	0.674
134	Phe	8.364	119.309	8.360	119.352	0.716	8.378	119.401	0.681			
135	Ile	8.033	120.291	8.044	120.367	0.921	8.047	120.506	0.743	8.056	120.582	0.825
136	Gly	7.822	105.793									
137	Leu	8.379	124.623	8.430	124.738	1.367	8.452	124.793	1.504	8.470	124.811	1.274
138	Val	8.588	119.093	8.605	119.160	0.799	8.614	119.231	0.798	8.618	119.285	0.678
139	Tyr	8.421	119.459	8.436	119.413	0.778	8.457	119.398	0.809	8.476	119.406	0.617
140	Tyr	8.156	120.630									
141	Leu	8.320	117.187	8.349	117.259	0.603	8.373	117.273	0.605	8.392	117.255	0.585
142	Val	8.497	115.607	8.488	115.966	1.139	8.475	116.057	1.272	8.457	116.083	0.918
143	Gly	8.114	111.690	8.159	111.519	0.881	8.200	111.393	0.762	8.226	111.323	0.756
145	Met	8.642	114.430	8.629	114.678	0.787	8.616	114.876	0.725	8.605	114.971	0.599
146	Thr	6.915	115.071	6.900	115.214	0.654	6.887	115.325	0.657	6.876	115.377	0.589
147	Glu	8.034	124.505	8.027	124.577	0.853	8.026	124.623	0.893	8.023	124.666	0.780
148	Ser	7.931	115.490	7.871	115.411	0.805	7.825	115.356	0.839	7.802	115.335	0.760
149	Ala	7.837	123.568	7.851	123.428	0.783	7.850	123.327	0.825	7.853	123.303	0.732
150	Ser	7.724	110.134	7.738	110.279	0.781	7.744	110.401	0.793	7.748	110.455	0.775
151	Gln	7.373	118.193	7.392	118.222	0.922	7.405	118.258	0.926	7.409	118.270	0.855
153	Ser	7.897	113.548	7.847	113.589	1.117	7.812	113.585	0.796	7.803	113.604	0.658
154	Ser	8.850	117.125	8.825	117.079	1.490	8.806	117.060	1.692	8.798	117.049	1.567
155	Gly	8.971	110.777	8.990	110.936	1.047	8.983	111.009	1.119	8.966	110.994	1.022
156	Ile	7.426	124.014	7.430	124.126	0.774	7.419	124.174	0.868	7.410	124.120	0.824
157	Lys	8.485	121.239	8.466	121.246	1.007	8.458	121.183	0.849	8.448	121.171	0.885
158	Ser	8.175	112.184	8.188	112.270	0.776	8.192	112.360	0.791	8.192	112.409	0.699

Table B.1: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0–0.728 X_{SDS} (0–0.8% SDS + 0.5% c7-DHPC)

#	Resi	0 X_{SDS}		0.400 X_{SDS} (0.2% SDS)			0.625 X_{SDS} (0.5% SDS)			0.728 X_{SDS} (0.8% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
159	Leu	7.539	123.172	7.509	123.334	0.783	7.500	123.382	0.787	7.487	123.425	0.630
160	Tyr	8.933	118.651	8.880	118.584	0.767	8.835	118.513	0.904	8.817	118.470	0.746
161	Val	8.465	118.036	8.475	117.975	0.740	8.478	117.907	0.703	8.473	117.868	0.525
163	Leu	8.059	117.674	8.072	117.591	0.604	8.057	117.338	1.033	8.045	117.226	0.884
164	Arg	9.424	124.793	9.400	124.714	0.777	9.364	124.602	0.876	9.343	124.513	0.657
165	Asn	8.533	119.461	8.520	119.551	0.778	8.508	119.639	0.760	8.500	119.667	0.612
167	Thr	7.686	114.269	7.671	114.241	0.768	7.670	114.180	0.767	7.656	114.164	0.687
168	Val	8.701	117.065	8.692	116.999	0.700	8.677	116.972	0.662	8.664	116.925	0.579
169	Val	7.325	112.502	7.342	112.556	0.797	7.347	112.634	0.768	7.345	112.658	0.662
171	Trp	9.736	119.386	9.758	119.394	0.864	9.763	119.371	0.746	9.767	119.340	0.686
173	Ile	7.672	113.796	7.665	113.723	0.737	7.659	113.672	0.868	7.642	113.653	0.625
176	Phe	7.363	115.864	7.353	116.013	0.814	7.333	116.086	1.010	7.336	116.143	1.278
177	Ile	7.705	120.178	7.724	120.271	0.745	7.736	120.213	1.087	7.735	120.218	0.926
178	Trp	7.999	123.281	8.013	123.549	0.685	8.003	123.714	0.760	8.000	123.783	0.579
179	Leu	7.988	113.817	7.972	113.731	0.741	7.953	113.733	0.551	7.947	113.737	0.918
180	Leu	8.257	113.389	8.323	113.368	0.684	8.364	113.382	0.775	8.381	113.377	0.567
181	Gly	7.895	110.691	7.898	110.665	0.802	7.896	110.609	0.733	7.891	110.581	0.606
184	Gly	6.995	105.808	6.976	105.637	0.792	6.960	105.528	0.723	6.950	105.475	0.578
185	Val	8.939	120.076	8.977	120.269	0.770	9.006	120.410	0.785	9.022	120.483	0.738
186	Ala	7.862	120.423	7.833	120.328	0.752	7.822	120.347	0.706	7.820	120.364	0.558
188	Leu	7.865	115.776	7.871	115.936	0.741	7.871	116.085	0.821	7.870	116.148	0.609
191	Thr	7.650	109.049	7.583	108.975	1.010	7.549	109.088	1.165	7.531	109.181	1.046
192	Val	7.093	121.746	7.105	121.939	0.761	7.109	122.050	0.729	7.107	122.081	0.632
193	Asp	8.076	119.864	8.073	119.847	0.588	8.069	119.822	0.517	8.065	119.848	0.511
194	Val	7.968	115.054	7.919	114.957	0.572						

Table B.1: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0–0.728 X_{SDS} (0–0.8% SDS + 0.5% c7-DHPC)

#	Resi	0 X_{SDS}		0.400 X_{SDS} (0.2% SDS)			0.625 X_{SDS} (0.5% SDS)			0.728 X_{SDS} (0.8% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
195	Ala	8.229	121.277	8.240	121.279	0.747	8.281	121.297	0.741	8.273	121.294	0.575
196	Leu	8.376	114.649	8.395	114.559	0.750	8.414	114.553	0.775	8.423	114.624	0.668
197	Ile	7.866	118.132	7.796	117.969	0.813	7.767	117.938	0.759	7.751	117.929	0.693
198	Val	8.545	122.256	8.496	122.210	0.778	8.483	122.162	0.766	8.479	122.123	0.699
199	Tyr	7.349	118.465	7.348	118.075	0.727	7.344	117.807	0.833	7.345	117.739	0.768
200	Leu	8.425	118.013	8.413	118.242	0.544	8.426	118.508	0.639	8.431	118.617	0.679
201	Asp	9.540	121.095	9.509	121.114	0.785	9.478	121.133	0.755	9.460	121.124	0.691
202	Leu	8.327	119.216	8.323	119.192	0.877						
203	Val	8.229	119.910	8.234	119.932	0.790	8.227	119.771	1.289	8.210	119.788	0.725
204	Thr	9.215	110.627	9.178	110.622	0.719	9.154	110.644	0.706	9.137	110.640	0.641
205	Lys	7.500	114.021	7.484	113.974	0.872	7.469	113.971	0.802	7.461	113.986	0.740
206	Val	7.678	121.105	7.671	121.037	0.670	7.660	121.022	0.688	7.668	121.058	0.679
207	Gly	8.949	110.058	8.984	110.111	0.534	8.999	110.130	0.643	8.993	110.135	0.567
208	Phe	7.314	121.292	7.360	121.386	0.728	7.375	121.455	0.731	7.374	121.503	0.580
209	Gly	7.912	103.927	7.931	104.039	0.813	7.947	104.094	0.907	7.960	104.143	0.823
210	Phe	9.021	121.627	9.031	121.633	0.611	9.025	121.616	0.606	9.021	121.611	0.489
211	Ile	8.337	121.320	8.340	121.532	1.035	8.356	121.561	0.983	8.357	121.532	1.865
212	Ala	7.633	122.212									
213	Leu	8.498	117.133	8.568	117.767	0.823	8.694	118.102	0.640	8.719	118.206	0.479
215	Ala	7.730	121.020	7.762	121.073	1.074	7.777	121.157	1.014	7.790	121.270	0.793
216	Ala	8.065	121.099	8.144	121.199	2.131	8.178	121.233	2.534	8.194	121.241	2.096
218	Thr	7.827	114.688	7.828	114.757	0.848	7.822	114.730	0.973	7.819	114.710	0.843
219	Leu	7.907	122.088	7.950	122.264	1.024	7.960	122.310	1.185	7.958	122.410	1.224
220	Arg	7.908	118.609	7.870	118.193	0.839	7.818	117.692	0.766	7.816	117.518	0.781
221	Ala	7.695	122.253									

Table B.1: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0–0.728 X_{SDS} (0–0.8% SDS + 0.5% c7-DHPC)

#	Resi	0 X_{SDS}		0.400 X_{SDS} (0.2% SDS)			0.625 X_{SDS} (0.5% SDS)			0.728 X_{SDS} (0.8% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
222	Glu	7.963	118.618	7.939	118.504	0.827	7.936	118.446	0.879	7.935	118.420	0.866
223	His	8.134	118.137	8.118	118.156	0.893	8.108	118.150	1.152	8.107	118.163	1.123
224	Gly	8.218	109.591	8.219	109.430	0.994	8.218	109.288	1.016	8.216	109.252	0.945
225	Glu	8.304	121.075	8.269	121.021	0.970	8.251	120.963	1.055	8.245	120.947	1.055
226	Ser	8.286	116.876	8.262	116.813	0.923	8.244	116.730	0.896	8.239	116.716	0.829
227	Leu	8.252	124.505	8.214	124.466	0.881	8.192	124.446	0.928	8.185	124.451	0.881
228	Ala	8.106	124.210	8.087	124.230	0.882	8.082	124.282	1.000	8.081	124.317	0.981
229	Gly	8.226	108.530	8.221	108.500	0.934	8.222	108.489	0.975	8.222	108.489	0.898
230	Val	7.839	118.836	7.823	118.829	0.915	7.811	118.832	0.881	7.804	118.835	0.815
231	Asp	8.402	123.994	8.398	123.988	0.901	8.393	123.955	0.876	8.390	123.961	0.792
232	Thr	8.006	114.002	8.002	114.026	0.872	7.989	114.024	0.843	7.982	114.042	0.776
233	Asp	8.316	123.032	8.316	123.002	0.879	8.315	122.969	0.834	8.314	122.982	0.758
234	Thr	8.001	117.048	7.994	117.004	0.898	7.988	116.982	0.872	7.988	116.999	0.798
236	Ala	8.296	124.685	8.292	124.646	0.859	8.296	124.757	0.805	8.298	124.857	0.708
237	Val	7.930	119.238	7.924	119.177	0.929	7.943	119.228	0.865	7.955	119.273	0.772
238	Ala	8.216	127.141	8.207	127.053	0.887	8.196	127.023	0.761	8.183	126.980	0.642
239	Asp	8.113	119.558	8.108	119.525	0.933	8.105	119.562	0.858	8.099	119.582	0.721
240	Leu	8.057	122.523	8.055	122.501	0.883	8.042	122.479	0.784	8.027	122.480	0.655
241	Glu	8.155	119.722	8.149	119.699	0.872	8.134	120.014	0.761	8.126	120.284	0.643

Table B.2: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0.769–0.893 X_{SDS} (1.0–2.5% SDS)
Conditions: ~90 µM ^{15}N-pSRII, 50 mM sodium phosphate (pH 6.0), 50 mM NaCl, 0.5% c7-DHPC, 10% D_2O, 308 K.
Alternative conformations are labelled with letters.

Table B.2: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0.769–0.893 X_{SDS} (1.0–2.5% SDS + 0.5% c7-DHPC)

#	Resi	0.769 X_{SDS} (1.0% SDS)			0.834 X_{SDS} (1.5% SDS)			0.870 X_{SDS} (2.0% SDS)			0.893 X_{SDS} (2.5% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
2	Val	7.760	118.586	0.842	7.729	118.340	0.531	7.722	118.302	0.330	7.722	118.307	0.096
3	Gly	8.290	112.201	0.818	8.236	112.167	0.607	8.222	112.155	0.396			
5	Thr	8.177	107.633	0.710									
6	Thr	7.332	117.198	0.548									
7	Leu	7.241	119.553	0.466									
9	Trp	8.576	119.836	0.609	8.557	119.876	0.374						
11	Gly	7.921	105.088	0.742	7.913	105.063	0.453						
12	Ala	8.346	123.301	0.301	8.343	123.354	0.256						
13	Ile	8.841	116.557	0.477	8.831	116.575	0.210						
14	Gly	8.695	107.557	0.537	8.707	107.542	0.323	8.719	107.551	0.234			
15	Met	8.310	117.613	0.472									
16	Leu	8.811	125.361	0.516	8.814	125.411	0.262	8.811	125.399	0.200			
17	Val	8.531	119.009	0.518	8.538	119.004	0.266	8.537	119.015	0.206			
17b	Val	8.571	119.107	0.194	8.581	119.058	0.336	8.581	119.023	0.195	8.559	119.295	0.168
18	Gly	7.784	104.036	0.510	7.776	104.003	0.239						
21	Ala	8.045	121.345	0.608	8.049	121.339	0.374	8.073	121.456	0.317	8.087	121.466	0.532
22	Phe	8.645	118.191	0.563	8.648	118.162	0.292						
23	Ala	8.478	121.572	0.536	8.485	121.707	0.275	8.497	121.702	0.154	8.482	121.647	0.149
24	Trp	8.734	118.550	0.511	8.758	118.615	0.396						
25	Ala	8.766	121.220	0.612	8.783	121.274	0.380	8.782	121.291	0.214			

Table B.2: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0.769–0.893 X_{SDS} (1.0–2.5% SDS + 0.5% c7-DHPC)

#	Resi	0.769 X_{SDS} (1.0% SDS)			0.834 X_{SDS} (1.5% SDS)			0.870 X_{SDS} (2.0% SDS)			0.893 X_{SDS} (2.5% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
26	Gly	7.766	104.442	0.556	7.785	104.495	0.294						
27	Arg	7.185	120.861	0.595	7.194	120.913	0.357	7.181	120.917	0.260			
28	Asp	7.710	117.832	0.634	7.714	117.889	0.444	7.706	117.877	0.208			
29	Ala	7.170	123.063	0.744	7.177	123.089	0.528	7.176	123.091	0.328	7.179	123.099	0.106
30	Gly	8.719	109.837	0.780	8.707	109.854	0.565	8.695	109.835	0.318	8.688	109.805	0.091
31	Ser	8.593	115.919	1.452	8.576	115.918	1.061	8.570	115.910	0.646	8.566	115.917	0.186
32	Gly	8.797	112.969	1.430	8.779	113.257	1.031	8.772	113.269	0.648			
35	Arg	8.499	116.400	0.373	8.504	116.543	0.260	8.482	116.576	0.164	8.488	116.404	0.181
37	Tyr	7.221	116.596	0.400									
38	Val	8.490	117.325	0.485	8.471	117.320	0.276	8.446	117.298	0.166			
40	Leu	7.110	119.018	0.885	7.091	119.000	0.603	7.087	118.985	0.410	7.084	118.980	0.183
42	Gly	8.351	106.332	0.447	8.332	106.315	0.273						
44	Ser	7.885	109.806	0.532	7.885	109.770	0.338	7.890	109.883	0.259	7.883	109.910	0.337
45	Gly	8.856	114.893	0.669	8.856	114.969	0.370						
46	Ile	8.076	120.657	0.955	8.071	120.746	0.521						
47	Ala	6.684	120.695	0.554									
48	Ala	8.459	120.361	0.538	8.459	120.404	0.276						
49	Val	7.521	115.824	0.539	7.525	115.861	0.354	7.521	115.905	0.157			
50	Ala	8.412	122.496	0.574	8.410	122.481	0.361	8.407	122.464	0.191			
51	Tyr	8.793	115.565	0.489									
52	Ala	8.194	122.121	0.534	8.210	122.199	0.312	8.202	122.191	0.154	8.140	122.325	0.189
53	Val	8.345	114.373	0.539	8.364	114.386	0.304	8.362	114.432	0.183			
55	Ala	8.377	125.240	0.959	8.403	125.331	0.469	8.394	125.356	0.418			
56	Leu	7.344	116.184	0.815	7.369	116.290	0.294						
57	Gly	8.058	107.493	0.522	8.068	107.572	0.286						
58	Val	7.668	123.223	0.505	7.688	123.342	0.268	7.683	123.367	0.158			

Table B.2: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0.769–0.893 X_{SDS} (1.0–2.5% SDS + 0.5% c7-DHPC)

#	Resi	0.769 X_{SDS} (1.0% SDS)			0.834 X_{SDS} (1.5% SDS)			0.870 X_{SDS} (2.0% SDS)			0.893 X_{SDS} (2.5% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
59	Gly	8.330	111.473	0.544	8.289	111.390	0.427						
60	Trp	8.411	122.721	0.613	8.387	122.647	0.288				8.380	122.501	0.128
61	Val	8.992	127.490	0.721	8.992	127.561	0.379						
63	Val	8.352	126.417	0.699	8.338	126.445	0.393	8.334	126.489	0.220			
64	Ala	8.828	128.215	0.556	8.815	128.173	0.332	8.809	128.193	0.148			
66	Arg	8.673	121.163	0.668	8.685	121.132	0.285	8.681	121.154	0.190	8.684	121.089	0.122
67	Thr	8.013	118.172	0.771	8.057	118.197	0.440	8.069	118.196	0.543	8.058	117.971	0.176
68	Val	9.268	129.338	0.659	9.269	129.384	0.414						
69	Phe	8.024	127.407	1.163	8.017	127.446	0.725						
70	Val	8.519	123.347	0.563									
74	Ile	6.963	117.690	0.673	6.958	117.682	0.408						
75	Asp	7.568	117.541	0.638	7.561	117.550	0.304						
76	Trp	7.819	114.785	4.584	7.820	114.767	3.033	7.815	114.734	1.718	7.809	114.728	0.697
78	Leu	7.470	115.711	0.412	7.454	115.711	0.178	7.451	115.657	0.157			
79	Thr	9.781	107.135	0.498									
82	Leu	6.751	117.682	0.573									
83	Ile	7.832	122.951	0.485									
85	Tyr	9.045	122.793	0.664	9.061	122.858	0.308						
86	Phe	8.631	121.678	0.379									
87	Leu	8.557	116.989	0.356									
88	Gly	9.426	108.400	0.573	9.431	108.443	0.355						
89	Leu	9.056	126.842	0.529	9.079	126.916	0.367						
90	Leu	7.755	120.139	0.599	7.760	120.180	0.403	7.768	120.173	0.229			
91	Ala	8.185	117.249	0.510	8.199	117.299	0.295						
92	Gly	7.612	106.261	0.702	7.614	106.295	0.459	7.604	106.332	0.257			
93	Leu	7.260	117.993	0.222	7.261	117.972	0.252	7.263	117.998	0.131			

Table B.2: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0.769–0.893 X_{SDS} (1.0–2.5% SDS + 0.5% c7-DHPC)

#	Resi	0.769 X_{SDS} (1.0% SDS)			0.834 X_{SDS} (1.5% SDS)			0.870 X_{SDS} (2.0% SDS)			0.893 X_{SDS} (2.5% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
94	Asp	8.275	120.733	0.546									
95	Ser	8.383	114.161	0.664	8.367	114.128	0.431	8.364	114.109	0.271			
96	Arg	7.947	122.391	0.836	7.929	122.270	0.429	7.919	122.270	0.258	7.908	122.227	0.217
98	Phe	8.554	117.614	0.541	8.539	117.493	0.332	8.556	117.583	0.225			
98b	Phe	8.589	117.638	0.370	8.585	117.635	0.526	8.581	117.590	0.413	8.583	117.564	0.218
99	Gly	8.330	105.418	0.519	8.331	105.422	0.393						
100	Ile	7.996	123.351	0.411	8.017	123.269	0.321						
101	Val	7.732	116.253	0.652	7.719	116.299	0.319	7.715	116.209	0.202			
102	Ile	8.636	121.272	0.567	8.640	121.334	0.289						
103	Thr	7.934	122.662	0.995	7.958	122.699	0.499	7.964	122.694	0.647			
105	Asn	7.999	115.073	0.578	7.968	115.495	0.658	7.970	115.631	1.008	7.972	115.733	1.775
106	Thr	8.111	114.833	0.400	8.068	115.578	0.480	8.066	115.579	0.922	8.059	115.572	1.862
106b	Thr				8.109	115.710	0.361	8.102	115.732	0.747	8.095	115.733	1.941
108	Val	7.507	117.835	0.581	7.497	117.740	0.340	7.483	117.755	0.200			
109	Met	7.437	112.908	0.524	7.416	113.052	0.991	7.412	113.039	1.313	7.406	113.036	1.951
110	Leu	8.900	119.346	0.600	8.902	119.413	0.315						
111	Ala	8.499	121.411	0.574	8.502	121.383	0.313	8.497	121.406	0.134			
112	Gly	8.181	102.842	0.664	8.191	102.864	0.384						
113	Phe	7.951	122.013	1.876	7.905	122.046	2.350	7.889	122.015	3.364	7.873	121.989	5.126
114	Ala	8.748	119.604	0.515	8.758	119.670	0.286						
115	Gly	8.585	101.285	0.485	8.601	101.282	0.262						
116	Ala	7.807	121.479	1.036	7.816	121.418	0.686	7.830	121.435	0.333			
117	Met	7.167	113.336	0.499	7.180	113.477	0.291						
118	Val	7.214	121.360	0.497	7.219	121.327	0.217	7.216	121.286	0.155			
121	Ile	8.547	120.916	0.607	8.532	120.973	0.279	8.528	120.922	0.207			
122	Glu	9.711	120.958	0.446									

Table B.2: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0.769–0.893 X_{SDS} (1.0–2.5% SDS + 0.5% c7-DHPC)

#	Resi	0.769 X_{SDS} (1.0% SDS)			0.834 X_{SDS} (1.5% SDS)			0.870 X_{SDS} (2.0% SDS)			0.893 X_{SDS} (2.5% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
124	Tyr	8.089	118.768	0.836	8.101	118.719	0.514	8.093	118.693	0.329	8.093	118.665	0.154
125	Ala	7.810	121.985	0.660	7.797	122.288	1.120	7.797	122.233	1.578	7.796	122.179	2.262
126	Leu	6.888	118.270	0.424									
128	Gly	8.364	105.564	0.628									
129	Met	7.862	121.070	0.447	7.871	121.141	0.258	7.865	121.193	0.166			
131	Ala	8.524	124.277	0.596	8.518	124.302	0.308	8.510	124.287	0.161			
132	Val	7.928	116.891	0.359	7.950	116.776	0.493	7.948	116.779	0.499	7.952	116.817	0.586
133	Ala	7.807	121.029	0.489	7.818	121.017	0.348	7.816	120.981	0.218	7.836	120.871	0.143
135	Ile	8.051	120.617	0.703	8.056	120.722	0.465	8.063	120.732	0.261			
137	Leu	8.478	124.826	1.093									
138	Val	8.621	119.334	0.631	8.629	119.354	0.282						
139	Tyr	8.487	119.488	0.590	8.502	119.482	0.413	8.499	119.546	0.218			
141	Leu	8.404	117.258	0.475									
142	Val	8.444	116.080	0.756	8.422	116.071	0.443	8.419	116.101	0.269			
143	Gly	8.235	111.295	0.693	8.270	111.247	0.655	8.263	111.258	0.548	8.259	111.290	0.681
145	Met	8.608	115.026	0.590									
146	Thr	6.872	115.420	0.531	6.862	115.433	0.461	6.857	115.464	0.326			
147	Glu	8.024	124.681	0.850	8.016	124.738	0.571						
148	Ser	7.792	115.325	0.730	7.782	115.314	0.495	7.772	115.280	0.295			
149	Ala	7.853	123.308	0.659	7.853	123.302	0.443	7.850	123.253	0.307			
150	Ser	7.750	110.485	0.738	7.755	110.521	0.503	7.751	110.541	0.291			
151	Gln	7.410	118.279	0.805	7.413	118.295	0.496	7.413	118.250	0.263			
153	Ser	7.800	113.603	0.641	7.794	113.601	0.410	7.789	113.605	0.212	7.777	113.589	0.110
154	Ser	8.794	117.055	1.484	8.790	117.079	0.915	8.786	117.089	0.557			
155	Gly	8.954	110.968	0.864	8.934	110.951	0.413	8.921	110.875	0.226			
156	Ile	7.404	124.089	0.784	7.394	124.023	0.602	7.388	124.010	0.255			

Table B.2: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0.769–0.893 X_{SDS} (1.0–2.5% SDS + 0.5% c7-DHPC)

#	Resi	0.769 X_{SDS} (1.0% SDS)			0.834 X_{SDS} (1.5% SDS)			0.870 X_{SDS} (2.0% SDS)			0.893 X_{SDS} (2.5% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
157	Lys	8.443	121.165	0.937	8.436	121.167	0.610	8.431	121.184	0.445	8.423	121.224	0.228
158	Ser	8.192	112.433	0.694	8.192	112.488	0.478	8.188	112.488	0.274			
159	Leu	7.481	123.450	0.561	7.469	123.513	0.310	7.469	123.597	0.165			
160	Tyr	8.810	118.457	0.724	8.795	118.418	0.544	8.791	118.441	0.312			
161	Val	8.471	117.841	0.491				8.473	117.814	0.185			
163	Leu	8.045	117.198	0.564	8.052	117.400	1.939	8.045	117.284	2.681	8.044	117.192	4.210
164	Arg	9.338	124.481	0.527									
165	Asn	8.496	119.618	0.606	8.484	119.582	0.334	8.484	119.574	0.262			
167	Thr	7.660	114.130	0.612									
168	Val	8.656	116.885	0.502	8.646	116.790	0.278	8.636	116.782	0.191			
169	Val	7.339	112.696	0.552	7.334	112.700	0.402						
171	Trp	9.761	119.343	0.576				9.756	119.320	0.310			
173	Ile	7.640	113.617	0.687									
176	Phe	7.338	116.152	1.106	7.322	116.152	0.469	7.316	116.210	0.236			
177	Ile	7.738	120.233	0.992	7.738	120.222	0.502						
178	Trp	8.002	123.818	1.260	7.960	123.466	3.488	7.944	123.377	5.374	7.928	123.269	8.118
179	Leu	7.950	113.627	0.419									
180	Leu	8.389	113.397	0.533	8.409	113.426	0.364						
181	Gly	7.885	110.566	0.494	7.877	110.543	0.276						
184	Gly	6.943	105.437	0.504	6.933	105.410	0.305						
185	Val	9.035	120.557	0.674	9.051	120.612	0.383						
186	Ala	7.819	120.378	0.520	7.819	120.374	0.270	7.812	120.361	0.178	7.784	120.134	0.167
188	Leu	7.876	116.190	0.563	7.882	116.219	0.437	7.886	116.317	0.292	7.881	116.285	0.287
191	Thr	7.520	109.217	1.014	7.508	109.346	0.498	7.506	109.383	0.307			
192	Val	7.103	122.084	0.593	7.096	122.040	0.270						
193	Asp	8.060	119.842	0.611	8.050	119.790	0.428						

Table B.2: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0.769–0.893 X_{SDS} (1.0–2.5% SDS + 0.5% c7-DHPC)

#	Resi	0.769 X_{SDS} (1.0% SDS)			0.834 X_{SDS} (1.5% SDS)			0.870 X_{SDS} (2.0% SDS)			0.893 X_{SDS} (2.5% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
195	Ala	8.273	121.273	0.519	8.272	121.262	0.277	8.262	121.290	0.171			
196	Leu	8.426	114.650	0.587	8.427	114.665	0.350						
197	Ile	7.750	117.943	0.542									
198	Val	8.477	122.099	0.531	8.462	122.090	0.321						
199	Tyr	7.338	117.742	0.705	7.331	117.694	0.415	7.331	117.734	0.270			
200	Leu	8.433	118.666	0.491									
201	Asp	9.456	121.129	0.571	9.444	121.120	0.283						
203	Val	8.209	119.764	0.585	8.200	119.722	0.375				8.175	119.758	0.265
204	Thr	9.129	110.644	0.615	9.126	110.656	0.373						
205	Lys	7.461	114.004	0.538	7.450	113.976	0.313						
206	Val	7.672	121.041	0.753	7.697	121.348	0.636						
207	Gly	8.986	110.152	0.622									
208	Phe	7.375	121.518	0.564									
209	Gly	7.968	104.163	0.547	7.979	104.208	0.457						
210	Phe	9.022	121.622	0.458	9.028	121.640	0.251						
211	Ile	8.368	121.585	2.126	8.379	121.667	1.016	8.383	121.689	0.828	8.421	121.721	0.386
213	Leu	8.722	118.215	0.475	8.731	118.211	0.272						
215	Ala	7.789	121.307	0.615									
216	Ala	8.203	121.231	1.847	8.216	121.203	1.093	8.219	121.180	0.729	8.232	121.219	0.458
218	Thr	7.813	114.689	0.684	7.807	114.597	0.330	7.801	114.548	0.202	7.807	114.589	0.258
219	Leu	7.961	122.286	1.224	7.960	122.201	0.555	7.942	122.220	0.410	7.931	122.188	0.412
220	Arg	7.817	117.467	0.789	7.825	117.400	0.577	7.819	117.377	0.363	7.822	117.400	0.115
222	Glu	7.935	118.413	0.842	7.937	118.406	0.602	7.935	118.387	0.331			
222b	Glu							7.910	118.685	0.103	7.913	118.666	0.204
223	His	8.107	118.176	1.079	8.110	118.206	0.788	8.108	118.196	0.503	8.099	118.212	0.359
224	Gly	8.214	109.248	0.913	8.211	109.258	0.696	8.204	109.256	0.445	8.186	109.230	0.489

Table B.2: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0.769–0.893 X_{SDS} (1.0–2.5% SDS + 0.5% c7-DHPC)

#	Resi	0.769 X_{SDS} (1.0% SDS)			0.834 X_{SDS} (1.5% SDS)			0.870 X_{SDS} (2.0% SDS)			0.893 X_{SDS} (2.5% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
225	Glu	8.242	120.943	1.023	8.240	120.944	0.733	8.236	120.939	0.465	8.235	120.945	0.126
225b	Glu				8.218	120.635	0.205	8.213	120.585	0.325	8.211	120.545	0.492
226	Ser	8.238	116.716	0.780	8.237	116.726	0.547	8.234	116.730	0.335	8.231	116.743	0.098
227	Leu	8.182	124.456	0.869	8.179	124.461	0.633	8.177	124.463	0.391	8.174	124.458	0.117
227b	Leu										8.186	123.998	0.017
228	Ala	8.080	124.330	0.966	8.078	124.349	0.706	8.074	124.351	0.445	8.070	124.353	0.129
229	Gly	8.220	108.488	0.863	8.218	108.477	0.608	8.213	108.471	0.373	8.208	108.460	0.111
230	Val	7.802	118.841	0.776	7.799	118.857	0.550	7.795	118.860	0.336	7.790	118.870	0.106
231	Asp	8.388	123.970	0.765	8.382	124.000	0.591	8.372	123.993	0.431	8.357	123.997	0.489
232	Thr	7.980	114.058	0.735	7.978	114.096	0.521	7.976	114.115	0.319	7.971	114.150	0.101
233	Asp	8.313	123.001	0.714	8.312	123.050	0.494	8.308	123.066	0.292	8.306	123.095	0.070
233b	Asp				8.306	122.683	0.145	8.302	122.688	0.261	8.298	122.702	0.411
234	Thr	7.988	117.008	0.753	7.991	117.024	0.534	7.990	117.027	0.322	7.990	117.019	0.089
236	Ala	8.297	124.892	0.665	8.296	124.934	0.459	8.293	124.934	0.275	8.288	124.911	0.075
237	Val	7.956	119.269	0.732	7.957	119.216	0.493	7.952	119.182	0.303	7.948	119.154	0.098
238	Ala	8.173	126.914	0.589	8.150	126.712	0.388	8.137	126.613	0.228	8.126	126.503	0.061
239	Asp	8.093	119.574	0.641	8.076	119.519	0.393	8.067	119.475	0.229	8.057	119.434	0.065
240	Leu	8.017	122.476	0.615	7.997	122.455	0.417	7.987	122.442	0.256	7.977	122.597	0.078
241	Glu	8.122	120.386	0.604	8.119	120.516	0.414	8.112	120.497	0.245	8.109	120.486	0.062

Table B.3: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0.909–0.990 X_{SDS} (3–30% SDS)
Conditions: ~90 µM ^{15}N-pSRII, 50 mM sodium phosphate (pH 6.0), 50 mM NaCl, 0.5% c7-DHPC, 10% D_2O, 308 K.
Alternative conformations are labelled with letters.

Table B.3: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0.909–0.990 X_{SDS} (3–30% SDS + 0.5% c7-DHPC)

#	Resi	0.909 X_{SDS} (3% SDS)			0.971 X_{SDS} (10% SDS)			0.985 X_{SDS} (20% SDS)			0.990 X_{SDS} (30% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
2	Val	7.713	118.371	0.109	7.710	118.409	0.134						
17b	Val	8.539	119.348	0.276	8.529	119.582	0.207						
21	Ala	8.083	121.468	0.921	8.060	121.556	0.737						
23	Ala	8.448	121.678	0.193									
35	Arg	8.493	116.432	0.361	8.488	116.669	0.189						
38	Val	8.404	117.068	0.232	8.352	117.102	0.195						
40	Leu	7.085	118.848	0.355	7.039	118.802	0.286						
44	Ser	7.884	109.941	0.592									
52	Ala	8.134	122.317	0.586	8.085	122.384	0.326						
60	Trp	8.373	122.617	0.236	8.333	122.715	0.234						
67	Thr	8.062	117.958	0.512	8.026	118.001	0.350						
75	Asp	7.554	117.665	0.375	7.540	117.666	0.354						
98b	Phe	8.595	117.508	0.269	8.596	117.624	0.235						
101	Val	7.752	116.624	0.281	7.724	116.678	0.194						
105	Asn	7.972	115.851	2.490	7.959	115.992	1.412						
106	Thr	8.050	115.580	2.851	8.026	115.694	0.817						
106b	Thr	8.085	115.733	3.900	8.060	115.735	6.348	8.053	115.814	4.132	8.051	115.887	0.292
109	Met	7.401	113.041	2.366	7.376	113.079	2.066	7.363	113.127	0.832			

Table B.3: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0.909–0.990 X_{SDS} (3–30% SDS + 0.5% c7-DHPC)

#	Resi	0.909 X_{SDS} (3% SDS)			0.971 X_{SDS} (10% SDS)			0.985 X_{SDS} (20% SDS)			0.990 X_{SDS} (30% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
113	Phe	7.848	121.966	6.483	7.818	122.025	7.004	7.812	122.049	4.747	7.802	122.049	1.540
121	Ile	8.556	120.885	0.384	8.568	121.044	0.335						
124	Tyr	8.126	118.714	0.209									
125	Ala	7.795	122.110	2.733	7.773	122.112	1.374						
132	Val	7.947	116.793	0.497									
133	Ala	7.825	120.814	0.228	7.802	120.737	0.176						
143	Gly	8.247	111.319	1.067	8.216	111.462	0.382						
163	Leu	8.042	117.095	5.467	8.018	117.068	3.684	8.005	117.249	0.588			
178	Trp	7.905	123.131	10.408	7.883	123.187	11.888	7.898	123.366	10.343	7.898	123.313	5.686
186	Ala	7.780	120.112	0.330	7.749	120.080	0.146						
188	Leu				7.880	116.403	0.398						
203	Val	8.159	119.763	0.740	8.144	119.731	0.785						
211	Ile	8.427	121.642	0.571	8.426	121.653	0.482						
213	Leu	8.739	118.141	0.227	8.724	118.133	0.219						
216	Ala	8.237	121.315	0.357									
218	Thr	7.801	114.604	0.228	7.763	114.619	0.217	7.741	114.637	0.194	7.732	114.675	0.163
219	Leu	7.911	122.160	0.478	7.879	122.223	0.274	7.874	122.215	0.314			
220	Arg	7.746	117.447	0.126	7.721	117.440	0.157						
222b	Glu	7.915	118.683	0.301	7.892	118.688	0.129						
223	His	8.108	118.212	0.388	8.093	118.139	0.140						
224	Gly	8.183	109.235	0.595	8.161	109.269	0.505	8.147	109.323	0.208			
225b	Glu	8.211	120.496	0.639	8.172	120.544	0.668	8.132	120.683	0.401	8.114	120.716	0.080
226	Ser	8.230	116.792	0.017									

Table B.3: Backbone chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0.909–0.990 X_{SDS} (3–30% SDS + 0.5% c7-DHPC)

#	Resi	0.909 X_{SDS} (3% SDS)			0.971 X_{SDS} (10% SDS)			0.985 X_{SDS} (20% SDS)			0.990 X_{SDS} (30% SDS)		
		^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
227	Leu	8.163	124.451	0.017	8.177	124.415	0.017						
227b	Leu	8.181	124.040	0.049	8.165	124.029	0.035						
228	Ala	8.065	124.321	0.023	7.996	124.578	0.150	7.999	124.864	0.074	7.991	124.840	0.020
229	Gly	8.203	108.431	0.017									
230	Val	7.744	118.935	0.054	7.741	119.128	0.041						
231	Asp	8.349	124.057	0.559	8.323	124.256	0.523	8.311	124.344	0.399	8.304	124.398	0.254
232	Thr	7.952	114.155	0.040	7.938	114.317	0.022						
233b	Asp	8.292	122.724	0.510	8.267	122.959	0.550	8.258	123.181	0.422	8.254	123.288	0.284
234	Thr	8.030	116.914	0.100	7.951	116.949	0.020						
236	Ala	8.262	124.654	0.023	8.247	124.519	0.015						
237	Val	7.948	119.188	0.072	7.925	119.301	0.051						
239	Asp	7.896	119.123	0.037	7.828	118.656	0.080	7.818	118.596	0.059			
240	Leu	7.979	122.730	0.095	7.969	122.804	0.047						
241	Glu	8.028	120.440	0.033	8.001	120.586	0.044						

B.2 Tryptophan side-chain chemical shifts and intensity ratios for pSRII in SDS

Table B.4: Tryptophan side-chain chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0–0.728 X_{SDS} (0–0.8% SDS)
Conditions: ~90 μM ^{15}N-pSRII, 50 mM sodium phosphate (pH 6.0), 50 mM NaCl, 0.5% c7-DHPC, 10% D_2O, 308 K.

#	0 X_{SDS}		0.400 X_{SDS} (0.2% SDS)			0.625 X_{SDS} (0.5% SDS)			0.728 X_{SDS} (0.8% SDS)		
	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
9	10.223	128.543	10.159	128.354	0.845	10.086	128.103	0.864	10.034	127.922	0.752
24	10.249	129.921	10.143	129.610	1.039	10.032	129.289	1.125	9.966	129.089	0.988
60	10.097	127.529	10.154	127.730	0.779	10.183	127.867	0.800	10.199	127.946	0.715
76	9.560	127.531	9.560	127.610	0.820	9.553	127.627	0.815	9.546	127.630	0.668
171	10.077	128.464	10.063	128.461	0.849	10.051	128.418	0.701	10.043	128.409	0.625
178	9.357	125.573	9.334	125.341	0.792	9.327	125.218	0.873	9.326	125.153	0.674

Table B.5: Tryptophan side-chain chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0.769–0.893 X_{SDS} (1.0–2.5% SDS)
Conditions: ~90 µM ^{15}N-pSRII, 50 mM sodium phosphate (pH 6.0), 50 mM NaCl, 0.5% c7-DHPC, 10% D_2O, 308 K.
Alternative conformations are labelled with letters.

#	0.769 X_{SDS} (1.0% SDS)			0.834 X_{SDS} (1.5% SDS)			0.870 X_{SDS} (2.0% SDS)			0.893 X_{SDS} (2.5% SDS)		
	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
9	10.005	127.819	0.627	9.951	127.619	0.346	9.935	127.555	0.188			
9b							9.976	127.961	0.097	9.961	127.960	0.241
24	9.934	128.990	0.910	9.876	128.810	0.726	9.859	128.771	0.646	9.840	128.738	0.643
24b				9.903	129.394	0.102	9.882	129.405	0.238	9.867	129.388	0.499
60	10.206	127.989	0.517									
76	9.541	127.620	0.517	9.537	127.623	0.280	9.534	127.666	0.180			
171	10.038	128.388	0.588	10.025	128.449	0.312						
178	9.326	125.120	0.623	9.329	125.082	0.364	9.326	125.067	0.179			

Table B.6: Tryptophan side-chain chemical shifts and intensity ratios (relative to 0 X_{SDS}) of pSRII in 0.909–0.990 X_{SDS} (3–30% SDS)
Conditions: ~90 µM ^{15}N-pSRII, 50 mM sodium phosphate (pH 6.0), 50 mM NaCl, 0.5% c7-DHPC, 10% D_2O, 308 K.
Alternative conformations are labelled with letters.

#	0.909 X_{SDS} (3% SDS)			0.971 X_{SDS} (10% SDS)			0.985 X_{SDS} (20% SDS)			0.990 X_{SDS} (30% SDS)		
	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0	^{1}H (ppm)	^{15}N (ppm)	I/I_0
9b	9.940	127.951	0.491	9.880	127.945	0.691	9.867	128.071	0.187			
9c	9.747	127.779	0.110	9.706	127.751	1.123	9.682	127.724	0.329			
24	9.815	128.696	0.796	9.766	128.671	0.868	9.750	128.767	0.422			
24b	9.845	129.360	0.712	9.796	129.302	1.055	9.776	129.297	0.210			
24c	9.557	128.773	0.118	9.514	128.682	0.281						

B.3 H/D exchange intensity ratios for pSRII

Table B.7: Intensity ratios for pSRII in deuterated vs. protonated solvent
Conditions: ~245 µM ^{15}N-pSRII, 50 mM sodium phosphate (pH 6.0), 50 mM NaCl, 0.5% c7-DHPC, 90% D_2O vs. 9% D_2O, 308 K.

Table B.7: Intensity ratios for pSRII in deuterated vs. protonated solvent

#	Resi	0 X_{SDS}	0.769 X_{SDS} (1.0% SDS)
2	Val	0.093	0.146
3	Gly	0.107	0.128
5	Thr	0.151	
7	Leu	0.622	0.759
9	Trp	0.553	0.661
10	Leu	0.538	
11	Gly	0.533	0.636
12	Ala	0.490	0.831
13	Ile	0.552	0.630
14	Gly	0.566	0.611
15	Met	0.481	0.569
16	Leu	0.570	0.636
17	Val	0.565	0.640
18	Gly	0.570	0.595
20	Leu	0.723	
21	Ala	0.604	0.719
22	Phe	0.612	0.700
23	Ala	0.603	0.649
24	Trp	0.601	0.733
25	Ala	0.693	0.820
26	Gly	0.262	0.163
27	Arg	0.150	0.123
28	Asp	0.126	0.137
29	Ala	0.129	0.155
30	Gly	0.068	0.109
31	Ser	0.095	0.100
32	Gly	0.090	0.112
34	Arg	0.134	
35	Arg	0.088	0.776
37	Tyr	0.589	0.408
38	Val	0.672	0.839
40	Leu		0.139
41	Val	0.632	
42	Gly	0.519	0.599
44	Ser	0.626	0.731
45	Gly	0.559	0.656

Table B.7: Intensity ratios for pSRII in deuterated vs. protonated solvent

#	Resi	0 X_{SDS}	0.769 X_{SDS} (1.0% SDS)
46	Ile	0.604	0.553
47	Ala	0.561	0.693
48	Ala	0.555	0.623
49	Val	0.545	0.648
50	Ala	0.606	0.707
51	Tyr	0.496	0.630
52	Ala	0.567	0.617
53	Val	0.592	0.638
54	Met	0.580	
55	Ala	0.420	0.680
56	Leu	0.508	
57	Gly	0.631	0.698
58	Val	0.569	0.653
60	Trp	0.115	0.313
61	Val	0.175	
63	Val	0.188	0.132
64	Ala	0.087	0.110
65	Glu	0.109	
66	Arg	0.133	
67	Thr		0.121
68	Val	0.239	
70	Val	0.477	0.508
74	Ile	0.525	0.601
75	Asp	0.507	0.471
76	Trp	0.549	
77	Ile	0.585	
78	Leu	0.743	0.559
79	Thr	0.610	0.607
80	Thr	0.537	0.807
82	Leu	0.523	
83	Ile	0.411	0.559
85	Tyr	0.523	0.614
86	Phe	0.560	0.575
87	Leu	0.471	0.662
88	Gly	0.516	0.637
89	Leu	0.541	0.638
90	Leu	0.513	0.610
91	Ala	0.750	0.856
92	Gly	0.461	0.271
93	Leu	0.135	0.188
94	Asp		0.155
95	Ser	0.095	0.099
96	Arg	0.107	0.177

Table B.7: Intensity ratios for pSRII in deuterated vs. protonated solvent

#	Resi	0 X_{SDS}	0.769 X_{SDS} (1.0% SDS)
97	Glu	0.119	
98	Phe	0.186	0.262
99	Gly	0.138	0.166
101	Val	0.641	0.599
102	Ile	0.606	
103	Thr	0.760	0.519
105	Asn	0.528	0.638
106	Thr	0.774	1.050
108	Val	0.524	0.543
109	Met	0.529	0.600
110	Leu	0.569	0.577
111	Ala	0.589	0.663
112	Gly	0.566	0.635
113	Phe	0.503	0.080
114	Ala	0.551	0.583
115	Gly	0.567	0.679
116	Ala	0.580	
117	Met	0.620	0.660
118	Val	0.666	0.741
120	Gly	0.099	
124	Tyr	0.284	0.255
125	Ala	0.268	0.139
126	Leu	0.875	1.199
127	Phe	0.882	
128	Gly	0.577	0.636
129	Met		0.149
130	Gly	0.561	
131	Ala	0.612	0.709
132	Val	0.821	
133	Ala	0.599	0.639
134	Phe	0.508	
135	Ile	0.501	0.600
136	Gly	0.591	
137	Leu	1.399	0.735
138	Val	0.546	0.613
139	Tyr	0.642	0.632
140	Tyr	0.637	
141	Leu	0.483	0.594
142	Val	0.638	0.119
143	Gly	0.468	0.503
145	Met	0.523	0.544
146	Thr	0.939	0.920
147	Glu	0.216	0.180

Table B.7: Intensity ratios for pSRII in deuterated vs. protonated solvent

#	Resi	0 X_{SDS}	0.769 X_{SDS} (1.0% SDS)
148	Ser	0.232	0.162
149	Ala	0.153	0.146
150	Ser	0.147	0.153
151	Gln	0.132	0.168
153	Ser	0.087	0.095
154	Ser		0.071
155	Gly	0.136	
156	Ile	0.136	0.170
157	Lys	0.162	0.164
158	Ser	0.109	0.141
159	Leu	0.107	0.115
160	Tyr	0.579	0.471
161	Val	0.597	0.727
163	Leu	0.768	0.734
164	Arg	0.606	0.755
165	Asn	0.537	0.609
167	Thr	0.893	0.999
168	Val	0.545	0.618
169	Val	0.563	0.684
171	Trp	0.508	0.572
173	Ile	0.499	0.460
174	Tyr	0.615	
176	Phe	0.514	
177	Ile	0.416	0.567
178	Trp	0.440	0.358
179	Leu		0.094
180	Leu	0.502	0.705
181	Gly	0.558	0.626
184	Gly	0.270	0.138
185	Val	0.566	0.639
186	Ala		0.709
188	Leu	0.532	0.641
191	Thr		0.115
192	Val	0.126	
194	Val	0.677	
195	Ala	0.642	0.682
196	Leu	0.476	0.517
197	Ile	0.437	0.409
198	Val	0.541	0.659
199	Tyr	0.484	0.650
200	Leu	0.489	0.633
201	Asp	0.531	0.675
202	Leu	0.608	

Table B.7: Intensity ratios for pSRII in deuterated vs. protonated solvent

#	Resi	0 X_{SDS}	0.769 X_{SDS} (1.0% SDS)
203	Val	0.506	0.520
204	Thr	0.536	0.765
205	Lys	0.470	0.503
206	Val	0.576	0.628
207	Gly	0.503	0.582
208	Phe	0.494	0.564
209	Gly	0.489	0.574
210	Phe	0.564	0.567
211	Ile	0.854	0.682
212	Ala	0.598	
213	Leu	0.714	0.847
214	Asp	0.334	
215	Ala	0.187	
216	Ala	0.108	0.245
218	Thr	0.131	
219	Leu	0.142	
220	Arg	0.116	0.219
221	Ala	0.121	
222	Glu	0.109	0.130
223	His	0.090	0.108
224	Gly	0.093	0.107
225	Glu	0.091	0.109
226	Ser	0.094	0.104
227	Leu	0.084	0.108
228	Ala	0.096	0.109
229	Gly	0.091	0.100
230	Val	0.105	0.114
231	Asp	0.098	0.107
232	Thr	0.095	0.106
233	Asp	0.093	0.102
234	Thr	0.096	0.126
236	Ala	0.100	0.113
237	Val	0.110	0.120
238	Ala	0.093	0.116
239	Asp	0.092	0.106
240	Leu	0.099	0.125
241	Glu	0.099	0.117
9	Trp-NεHε	0.098	0.081
24	Trp-NεHε	0.102	0.112
76	Trp-NεHε	0.255	
171	Trp-NεHε	0.589	0.788
178	Trp-NεHε	0.081	0.100

B.4 Chemical shifts for pSRII at acidic pHs

Table B.8: Backbone and tryptophan side-chain chemical shifts for pSRII at acidic pHs
Conditions: 112 µM ^{15}N-pSRII, 50 mM sodium phosphate (pH 2.0, 2.6, 3.7, 4.7), 50 mM NaCl, 0.5% c7-DHPC, 10% D_2O, 308 K.
Alternative conformations are labelled with letters.

Table B.8: Backbone and tryptophan side-chain chemical shifts for pSRII at acidic pHs

#	Resi	pH 4.7		pH 3.7		pH 2.6		pH 2.0	
		^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)
2	Val	7.963	120.278	7.962	120.288				
3	Gly	8.497	112.330	8.487	112.291				
5	Thr	8.352	108.276	8.358	108.331				
6	Thr	7.501	117.587	7.501	117.613				
7	Leu	7.370	120.109	7.389	120.155				
7b	Leu	7.379	120.162						
9	Trp	8.638	120.099	8.627	120.024				
10	Leu	8.230	120.203	8.298	120.051				
11	Gly	7.981	105.342						
12	Ala	8.392	123.382	8.414	123.563				
13	Ile	8.876	116.596	8.879	116.618				
14	Gly	8.686	107.740	8.682	107.718				
15	Met	8.375	117.952	8.390	118.148				
16	Leu	8.843	125.193	8.840	125.188				
17	Val	8.498	118.980	8.494	119.007				
18	Gly	7.795	104.247	7.780	104.175				

Table B.8: Backbone and tryptophan side-chain chemical shifts for pSRII at acidic pHs

#	Resi	pH 4.7		pH 3.7		pH 2.6		pH 2.0	
		^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)
21	Ala	8.002	121.296	7.988	121.254				
22	Phe	8.592	118.059	8.594	117.905				
23	Ala	8.431	121.345	8.457	121.385				
24	Trp	8.643	118.525	8.639	118.659				
25	Ala	8.647	120.923	8.658	120.873				
26	Gly	7.681	103.926	7.736	104.034				
27	Arg	7.207	120.845	7.257	120.956				
28	Asp	7.779	117.605	7.766	117.130				
29	Ala	7.160	123.196	7.173	123.122			7.226	122.931
30	Gly	8.668	109.370	8.530	108.944				
31	Ser	8.679	115.772	8.689	116.000	8.709	116.347	8.708	116.325
32	Gly	8.978	111.072	8.962	111.069				
35	Arg	8.505	115.923	8.590	116.039				
37	Tyr	7.226	116.970	7.261	117.012				
38	Val	8.639	117.950	8.655	118.015				
40	Leu	7.296	118.943	7.249	118.895	7.240	118.991	7.252	118.969
41	Val	8.158	118.276	8.199	118.405				
42	Gly	8.403	106.648	8.393	106.599				
44	Ser	7.858	109.758	7.868	109.743				
45	Gly	8.896	114.874	8.898	114.856				
47	Ala	6.695	120.876						
48	Ala	8.489	120.116	8.486	120.146				
49	Val	7.503	115.639	7.500	115.637				

Table B.8: Backbone and tryptophan side-chain chemical shifts for pSRII at acidic pHs

#	Resi	pH 4.7		pH 3.7		pH 2.6		pH 2.0	
		^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)
50	Ala	8.397	122.501	8.385	122.531				
51	Tyr	8.798	115.414						
52	Ala	8.133	121.698	8.130	121.696				
53	Val	8.261	113.775	8.263	113.736				
55	Ala	8.302	124.772	8.291	124.766				
56	Leu	7.258	115.996	7.254	115.996				
57	Gly	8.031	107.235						
58	Val	7.617	122.869	7.605	122.841				
59	Gly	8.806	112.232	8.817	112.349				
60	Trp	8.408	122.612	8.361	122.561				
61	Val	9.045	127.157	9.055	127.166				
63	Val	8.423	126.313	8.407	126.198				
64	Ala	8.843	128.397	8.864	128.330				
66	Arg	8.602	121.152	8.611	120.815				
68	Val	9.213	129.047	9.235	129.299				
77	Ile	8.257	120.302						
78	Leu	7.505	115.833						
80	Thr	8.184	117.457	8.182	117.253				
83	Ile	7.799	122.756						
85	Tyr	9.001	122.489						
86	Phe	8.561	121.585						
87	Leu	8.459	116.552						
88	Gly	9.362	108.174	9.360	108.197				

Table B.8: Backbone and tryptophan side-chain chemical shifts for pSRII at acidic pHs

#	Resi	pH 4.7		pH 3.7		pH 2.6		pH 2.0	
		^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)
89	Leu	8.906	126.058	8.920	126.107				
90	Leu	7.687	119.880	7.681	119.938				
91	Ala	8.102	116.704						
92	Gly	7.578	106.137	7.584	106.300				
93	Leu	7.252	117.952	7.258	117.903				
95	Ser	8.601	114.867	8.623	114.911				
96	Arg	8.146	123.351	8.133	123.281				
97	Glu	8.269	118.834	8.287	118.216				
98	Phe	8.575	118.461	8.578	118.526				
99	Gly	8.212	105.136	8.245	105.141				
100	Ile	7.745	123.497	7.750	123.498				
101	Val	7.713	116.305	7.723	116.433				
102	Ile	8.520	120.711	8.515	120.759	8.554	120.976	8.591	121.089
103	Thr	7.951	122.583	7.947	122.626				
105	Asn	8.056	115.282	8.058	115.276				
106	Thr	8.061	114.560	8.063	114.543				
108	Val	7.582	118.063	7.572	118.068				
110	Leu	8.811	118.700	8.792	118.598				
111	Ala	8.524	121.356	8.528	121.367				
112	Gly	8.105	102.888						
113	Phe	7.975	121.908	7.976	121.812				
114	Ala	8.706	119.178	8.682	119.063				
115	Gly	8.527	101.597	8.523	101.851				

Table B.8: Backbone and tryptophan side-chain chemical shifts for pSRII at acidic pHs

#	Resi	pH 4.7		pH 3.7		pH 2.6		pH 2.0	
		^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)
116	Ala	7.857	121.712	7.961	122.214				
117	Met	7.138	112.831	7.161	112.909				
118	Val	7.236	121.157	7.257	120.956				
120	Gly	7.848	105.650	7.817	105.575				
121	Ile	8.588	120.266	8.597	120.328	8.618	120.546	8.617	120.536
122	Glu	9.867	121.949	9.346	121.166				
125	Ala	7.813	122.356						
126	Leu	7.033	118.599	7.077	118.677				
128	Gly	8.424	105.738	8.442	105.769				
129	Met	7.862	120.599	7.859	120.601				
130	Gly	7.921	106.039	7.945	106.135				
131	Ala	8.566	124.240	8.576	124.262				
132	Val	7.931	117.117	7.945	117.186	7.950	117.128	7.941	117.115
133	Ala	7.809	121.076	7.824	121.124				
134	Phe	8.378	119.243	8.386	119.281				
135	Ile	8.012	120.229	8.012	120.255				
136	Gly	7.822	105.794	7.829	105.784				
137	Leu	8.389	124.669	8.391	124.707				
138	Val	8.560	119.105	8.501	119.118				
139	Tyr	8.403	119.481	8.403	119.470				
140	Tyr	8.154	120.407	8.191	120.131				
141	Leu	8.317	117.460	8.323	117.532				
142	Val	8.482	115.732	8.485	115.815				

Table B.8: Backbone and tryptophan side-chain chemical shifts for pSRII at acidic pHs

#	Resi	pH 4.7		pH 3.7		pH 2.6		pH 2.0	
		^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)
143	Gly	8.121	111.589	8.129	111.444				
145	Met	8.683	114.520	8.694	114.545				
146	Thr	6.913	115.214	6.897	115.244				
147	Glu	8.069	124.383	8.072	123.880				
148	Ser	7.873	115.235	7.850	115.227				
149	Ala	7.823	123.434	7.816	123.399				
150	Ser	7.728	110.223	7.724	110.254				
151	Gln	7.368	118.187	7.359	118.140				
153	Ser	7.914	113.511	7.927	113.503				
154	Ser	8.850	117.208	8.841	117.254				
155	Gly	8.882	110.525	8.815	110.352				
156	Ile	7.377	123.697	7.343	123.460				
157	Lys	8.482	121.241	8.475	121.216				
158	Ser	8.178	112.217	8.179	112.257			8.184	112.427
159	Leu	7.522	123.195	7.521	123.238				
160	Tyr	8.922	118.662	8.910	118.645				
161	Val	8.477	118.067	8.484	118.121				
163	Leu	8.046	117.609	8.045	117.606				
164	Arg	9.421	124.834	9.419	124.863				
165	Asn	8.533	119.440	8.542	119.424	8.565	119.874	8.555	119.886
165b	Asn					8.531	119.718	8.519	119.734
167	Thr	7.690	114.295	7.691	114.373				
168	Val	8.685	117.152	8.678	117.208				

Table B.8: Backbone and tryptophan side-chain chemical shifts for pSRII at acidic pHs

#	Resi	pH 4.7		pH 3.7		pH 2.6		pH 2.0	
		^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)
169	Val	7.344	112.549	7.352	112.586				
171	Trp	9.721	119.386	9.708	119.327				
176	Phe	7.320	115.729	7.274	115.517				
177	Ile	7.699	119.815	7.714	119.560				
180	Leu	8.197	113.492	8.148	113.585				
181	Gly	7.868	110.429						
184	Gly	6.977	105.754	6.959	105.693				
185	Val	8.932	119.957	8.923	119.818				
186	Ala	7.878	120.508	7.904	120.539				
188	Leu	7.879	115.889						
191	Thr	7.681	108.971	7.705	108.924				
192	Val	7.095	121.508	7.090	121.306				
198	Val	8.586	122.189	8.617	122.127				
201	Asp	9.526	121.112	9.506	121.147				
202	Leu	8.331	119.211	8.342	119.190				
203	Val	8.238	119.967	8.264	120.132				
204	Thr	9.225	110.680						
205	Lys	7.472	114.074						
208	Phe	7.345	121.306						
209	Gly	7.925	103.904	7.932	103.829				
210	Phe	9.036	121.666						
212	Ala	7.630	122.106	7.625	122.027				
213	Leu	8.556	117.312	8.587	117.387				

Table B.8: Backbone and tryptophan side-chain chemical shifts for pSRII at acidic pHs

#	Resi	pH 4.7		pH 3.7		pH 2.6		pH 2.0	
		^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)
218	Thr	7.847	114.219	7.860	113.871	7.902	113.434	7.891	113.432
219	Leu	7.871	122.194	7.833	122.286			7.785	122.447
221	Ala	7.744	122.446	7.793	122.599	7.861	122.893	7.855	122.840
221b	Ala					7.864	122.130	7.846	122.031
222	Glu	7.957	118.448	7.954	118.263	7.962	118.719	7.950	118.783
222b	Glu					7.931	118.168	7.919	118.165
224	Gly	8.289	109.597	8.302	109.567	8.326	109.662	8.327	109.660
224b	Gly					8.230	109.087	8.231	109.087
225	Glu	8.282	120.745	8.222	120.394	8.147	120.038	8.141	120.009
225b	Glu					8.087	119.619	8.081	119.612
226	Ser	8.283	116.840	8.277	116.848	8.284	116.982	8.283	117.013
226b	Ser					8.222	116.417	8.221	116.429
227	Leu	8.262	124.448	8.257	124.465	8.253	124.498	8.253	124.540
227b	Leu					8.285	124.102	8.288	124.107
228	Ala	8.110	124.135	8.107	124.139	8.111	124.256	8.110	124.280
228b	Ala					8.057	123.444	8.058	123.440
229	Gly	8.221	108.460	8.214	108.384	8.214	108.256	8.215	108.244
229b	Gly					8.189	107.726	8.189	107.711
230	Val	7.819	118.812	7.805	118.849	7.810	119.002	7.812	119.010
230b	Val					7.766	118.753	7.767	118.770
231	Asp	8.424	123.678	8.451	123.301	8.516	122.473	8.523	122.387
231b	Asp					8.473	121.924	8.477	121.819
232	Thr	8.018	114.083	8.030	114.236	8.027	114.388	8.022	114.391

Table B.8: Backbone and tryptophan side-chain chemical shifts for pSRII at acidic pHs

#	Resi	pH 4.7		pH 3.7		pH 2.6		pH 2.0	
		^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)	^{1}H (ppm)	^{15}N (ppm)
232b	Thr					7.979	113.882	7.974	113.846
233	Asp	8.318	122.696	8.319	122.269	8.325	121.272	8.325	121.174
233b	Asp					8.294	121.000	8.291	120.896
234	Thr	7.989	116.834	7.980	116.778	7.970	116.774	7.966	116.767
234b	Thr					7.914	116.427	7.908	116.396
236	Ala	8.272	124.289	8.257	124.068	8.213	123.769	8.204	123.661
236b	Ala					8.190	122.886	8.180	122.770
237	Val	7.891	118.925	7.885	118.946	7.878	118.873	7.872	118.825
237b	Val					7.818	118.279	7.809	118.213
238	Ala	8.194	126.664	8.190	126.504	8.192	126.213	8.191	126.159
238b	Ala					8.154	125.401	8.150	125.297
239	Asp	8.107	119.257	8.117	118.885				
240	Leu	8.062	122.528	8.034	122.453	7.946	122.095	7.927	121.972
240b	Leu					7.893	121.594	7.877	121.487
241	Glu	8.113	118.876	8.067	118.628				
9	Trp-NεHε	10.210	128.515	10.209	128.550				
24	Trp-NεHε	10.243	129.902	10.242	129.889	10.237	129.916	10.248	129.935
24b	Trp-NεHε							10.199	129.542
60	Trp-NεHε	10.129	127.505						
76	Trp-NεHε	9.558	127.435						
171	Trp-NεHε	10.067	128.394						
178	Trp-NεHε	9.403	126.001						

Appendix

C Scalar coupling

C.1 RalAΔC-GDP

Table C.1: $^{1}J_{NH}$ of RalAΔC-GDP from IPAP-HSQC and SOFAST-IPAP-HSQC
Conditions: 700 µM ^{15}N-RalAΔC-GDP, 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM $MgCl_2$, 0.05% NaN_3, 10% D_2O, 298 K.
SNR (Add) and SNR (Sub) are the signal-to-noise ratios of peaks in the sub-spectrum obtained by taking the sum or the difference, respectively, of the in-phase (IP) and antiphase (AP) spectra.

Table C.1: $^{1}J_{NH}$ of RalAΔC-GDP from IPAP-HSQC and SOFAST-IPAP-HSQC

#	Resi	IPAP-HSQC									SOFAST-IPAP-HSQC					
		ns = 32 [A]			ns = 32 [B]			ns = 112			ns = 64 [A]			ns = 64 [B]		
		$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)
13	Ala	-92.860	100.769	56.503	-92.442	76.444	46.289	-92.829	163.708	94.431	-92.276	120.864	66.308	-92.787	102.445	60.079
14	Leu	-91.768	91.152	35.941	-92.119	73.048	27.348	-91.875	152.417	58.954	-92.146	103.593	44.664	-93.270	93.373	40.111
15	His	-94.301	78.324	43.301	-94.050	64.031	37.606	-94.174	132.660	73.413	-94.010	99.206	62.848	-94.112	86.684	56.813
17	Val	-93.830	73.441	40.398	-93.606	69.639	38.958	-93.748	137.364	80.106	-93.791	115.042	62.126	-93.527	104.436	56.860
18	Ile	-93.339	88.857	43.126	-93.741	87.612	45.698	-93.692	167.471	87.324	-93.552	135.947	68.839	-93.833	123.165	63.900
19	Met	-93.440	70.241	35.530	-93.362	63.644	32.779	-93.524	128.896	67.222	-93.555	102.996	51.688	-93.534	94.336	47.586
20	Val	-95.875	117.603	74.342	-96.093	105.973	71.612	-96.213	212.632	142.053	-97.819	170.656	110.737	-96.686	162.766	104.276
21	Gly	-93.988	100.561	52.512	-94.086	101.682	50.404	-94.033	189.111	95.812	-94.230	89.231	45.537	-94.114	88.893	47.099
23	Gly	-94.710	137.271	72.859	-94.838	132.074	71.057	-94.874	250.266	139.623	-94.812	187.380	100.618	-94.606	180.599	99.565
24	Gly	-93.302	56.613	30.985	-93.146	56.001	31.198	-93.496	101.612	52.325				-92.969	28.548	15.126
26	Gly	-93.716	92.354	45.915	-93.421	88.668	44.207	-93.474	168.412	84.941	-93.425	132.134	67.448	-93.738	125.972	61.673

Table C.1: $^{1}J_{NH}$ of RalAΔC-GDP from IPAP-HSQC and SOFAST-IPAP-HSQC

#	Resi	IPAP-HSQC									SOFAST-IPAP-HSQC					
		ns = 32 [A]			ns = 32 [B]			ns = 112			ns = 64 [A]			ns = 64 [B]		
		$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)
27	Lys	-90.250	46.519	23.128	-91.400	44.013	23.114	-90.919	83.171	46.717						
28	Ser	-94.150	102.436	50.287	-94.461	92.980	49.715	-94.220	186.288	98.891	-94.485	144.013	72.725	-94.355	140.431	70.510
29	Ala	-93.640	82.544	45.013	-93.887	77.106	44.496	-94.095	158.063	88.969	-93.978	119.986	66.215	-94.238	111.528	60.880
30	Leu	-94.821	120.911	62.603	-94.397	116.149	58.770	-94.634	225.804	119.696	-94.735	181.152	92.007	-94.779	169.977	86.791
32	Leu	-93.115	87.815	43.581	-92.978	83.012	43.388	-93.107	161.826	88.744	-93.222	120.630	63.203	-93.367	117.493	61.229
35	Met	-92.849	98.074	51.788	-93.031	93.535	48.481	-92.998	182.525	93.493	-93.576	146.122	69.759	-93.411	139.799	66.871
36	Tyr	-90.351	106.705	59.405	-90.474	105.763	59.473	-90.438	203.223	112.198	-90.538	166.533	88.157	-89.988	156.907	82.314
39	Phe	-92.553	135.152	69.495	-92.589	125.627	67.839	-92.573	251.207	134.083	-92.409	189.437	99.407	-92.558	178.045	91.215
40	Val	-92.933	33.823	17.866	-91.721	31.570	16.556									
41	Glu	-92.780	111.985	55.159	-92.749	108.899	55.070	-92.925	210.750	106.325	-92.666	156.895	83.784	-92.812	148.847	75.949
43	Tyr	-94.312	50.212	26.051	-94.239	43.011	24.745	-94.089	87.405	49.542	-94.836	60.796	33.398	-94.409	60.726	32.412
44	Glu	-93.225	96.469	53.309	-93.121	89.465	52.536	-93.301	175.939	104.464	-93.044	132.775	75.853	-93.269	123.190	69.700
47	Lys	-93.068	60.479	33.152	-93.313	59.965	32.801	-92.710	116.665	61.532	-92.783	67.564	33.832	-91.815	70.320	33.264
48	Ala	-92.485	42.777	25.113	-93.038	36.127	23.682	-92.791	77.432	48.755	-92.250	56.106	34.898	-92.290	51.742	29.967
50	Ser	-91.696	133.270	69.541	-91.857	125.625	70.218	-91.894	246.502	134.444	-91.768	175.672	102.288	-91.736	166.670	94.197
51	Tyr	-92.969	82.761	47.865	-92.901	78.893	44.944	-92.726	157.122	88.487	-93.059	128.947	70.114	-92.877	122.159	67.381
52	Arg	-93.134	157.658	80.744	-93.427	152.050	80.674	-93.221	295.426	156.459	-93.436	230.260	120.931	-93.365	218.233	112.412
54	Lys	-93.037	161.435	86.489	-92.741	154.284	87.197	-92.849	306.717	169.406	-92.922	228.927	127.435	-93.052	216.748	118.735
56	Val	-92.498	205.923	106.258	-92.890	204.389	107.277	-92.765	393.275	210.755	-92.713	291.514	153.064	-92.988	281.839	149.114
57	Leu	-92.941	130.150	62.560	-93.188	125.107	61.684	-93.160	248.384	118.986	-93.070	195.807	94.593	-93.223	183.786	90.195
58	Asp	-93.115	41.898	21.363	-93.833	36.195	20.439	-94.294	71.881	39.806	-93.325	56.450	28.967	-94.042	49.620	26.029
59	Gly	-92.565	111.058	63.279	-92.509	106.946	63.309	-92.569	200.401	115.059	-92.449	149.681	85.479	-92.459	146.992	81.444
60	Glu	-95.202	176.639	98.393												
61	Glu	-93.310	151.975	81.111	-93.552	126.034	69.629	-93.498	262.497	149.543	-93.471	186.507	105.869	-93.442	164.568	90.606

Table C.1: $^{1}J_{NH}$ of RalAΔC-GDP from IPAP-HSQC and SOFAST-IPAP-HSQC

#	Resi	IPAP-HSQC									SOFAST-IPAP-HSQC					
		ns = 32 [A]			ns = 32 [B]			ns = 112			ns = 64 [A]			ns = 64 [B]		
		$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)
62	Val	-93.667	95.790	51.250	-94.182	76.068	43.003	-93.769	159.003	91.181	-93.679	123.092	66.063	-93.371	108.070	60.315
63	Gln	-92.367	90.300	54.181	-91.987	70.971	44.647	-92.751	152.417	93.527	-92.085	109.725	64.695	-92.034	99.341	61.393
64	Ile	-92.835	82.584	43.918	-93.110	68.004	41.282	-93.202	143.009	79.720	-93.236	107.548	59.693	-93.408	102.771	57.766
65	Asp	-92.747	90.721	43.552	-92.898	80.026	40.984	-93.364	162.767	81.148	-92.884	130.631	65.839	-92.729	118.889	59.744
66	Ile	-93.547	87.141	43.568	-93.238	80.436	43.364	-93.451	159.944	87.867	-93.486	133.931	71.656	-93.360	127.881	70.221
67	Leu	-93.993	90.177	40.658	-93.593	86.083	42.464	-93.604	169.353	82.837	-93.674	135.707	66.175	-93.667	128.611	60.431
68	Asp	-93.936	67.698	33.144	-94.614	63.676	32.028	-93.892	126.074	65.369	-93.700	99.723	48.962	-93.433	92.009	47.139
71	Gly	-94.004	76.273	53.786	-94.151	63.960	47.780	-93.970	127.015	93.744	-94.035	90.745	63.725	-94.191	83.754	57.995
72	Gln	-92.870	59.244	43.240	-92.650	49.500	38.149	-92.349	103.493	77.728	-92.378	71.058	50.607	-92.223	64.724	44.162
73	Glu	-92.548	60.682	41.943	-93.300	49.188	36.203	-92.319	105.375	75.099	-92.563	67.545	49.223	-92.166	61.915	45.801
74	Asp	-93.388	144.312	97.566	-93.313	135.489	96.404	-93.253	270.964	192.807	-93.387	196.043	134.865	-93.169	179.112	122.935
78	Ile	-92.002	81.400	44.064	-91.952	77.284	40.825	-92.080	148.654	82.585	-92.059	104.340	60.476	-92.384	96.892	55.850
79	Arg	-92.191	50.671	27.927	-92.610	48.148	25.019	-93.107	93.332	50.055	-92.423	69.388	38.321	-92.924	62.795	32.720
80	Asp	-93.524	115.503	67.358	-93.722	103.513	61.432	-93.806	207.928	120.010	-93.743	151.438	85.698	-93.766	140.741	76.368
83	Phe	-94.553	77.277	46.289	-94.843	70.978	44.185	-93.992	143.009	84.470	-94.052	110.385	64.490	-94.239	102.231	59.361
86	Gly	-96.199	118.028	60.043	-96.289	106.189	60.920	-96.241	204.164	110.936	-96.239	167.288	95.450	-96.157	155.437	90.454
87	Glu				-91.466	35.327	26.779				-90.477	63.813	46.163			
88	Gly	-94.277	105.296	61.732	-94.168	97.730	57.316	-94.232	183.465	111.233	-93.889	150.371	89.573	-93.945	148.134	86.651
89	Phe				-93.123	136.535	89.379	-93.998	247.443	162.380	-93.467	184.412	118.553			
91	Cys	-94.051	128.763	62.055	-94.151	123.496	62.666	-94.200	242.739	121.377	-94.227	187.146	96.701	-94.192	175.807	87.761
92	Val	-94.444	115.078	53.692	-94.351	109.052	54.535	-94.301	217.336	105.832	-94.336	177.798	85.882	-94.246	167.312	79.382
93	Phe	-94.869	124.369	61.065	-95.007	118.137	60.198	-94.468	234.271	118.021	-94.714	188.781	92.896	-94.581	183.954	90.836
94	Ser	-91.982	81.484	44.394	-92.500	77.405	42.175	-92.152	145.832	80.685	-92.096	111.528	60.412	-92.298	105.976	56.499
95	Ile	-95.220	97.451	49.070	-95.090	94.143	48.623	-95.214	180.643	96.104	-95.375	141.159	75.799	-95.646	136.304	69.364

Table C.1: $^{1}J_{NH}$ of RalAΔC-GDP from IPAP-HSQC and SOFAST-IPAP-HSQC

#	Resi	IPAP-HSQC									SOFAST-IPAP-HSQC					
		ns = 32 [A]			ns = 32 [B]			ns = 112			ns = 64 [A]			ns = 64 [B]		
		$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)
96	Thr	-92.978	121.049	56.957	-93.254	119.127	56.604	-93.230	232.390	110.817	-93.184	180.029	82.679	-93.199	170.060	80.310
97	Glu	-92.642	167.551	67.547	-93.046	161.733	72.081	-92.816	311.421	138.194	-92.992	237.769	106.336	-92.542	228.547	99.326
98	Met	-93.846	72.814	34.641	-93.035	66.140	37.086	-93.520	133.601	72.647	-93.863	105.498	54.161	-93.353	96.763	50.790
99	Glu	-92.624	85.880	45.262	-92.464	75.878	42.896	-92.933	152.417	84.813	-92.423	103.956	57.160	-93.100	99.950	54.088
100	Ser	-91.283	110.415	59.310	-91.904	109.752	60.878	-91.477	209.809	114.226	-91.591	163.461	88.196	-91.590	155.769	78.128
101	Phe	-92.767	93.423	41.111	-92.235	91.939	41.077	-92.945	176.880	79.915	-92.324	136.952	60.451	-92.484	127.587	54.888
102	Ala	-93.719	158.462	71.446	-93.676	154.890	73.201	-93.715	304.835	143.230	-93.918	227.059	103.328	-93.821	212.810	94.908
103	Ala	-92.877	169.041	79.923	-92.938	169.818	83.302	-92.929	319.889	158.654	-92.987	255.311	120.971	-93.054	238.956	112.373
104	Thr	-93.319	112.923	57.716	-93.100	106.694	56.065	-93.024	206.046	106.132	-93.196	161.288	80.213	-93.149	157.983	77.856
105	Ala	-93.385	180.790	81.492	-93.525	180.624	82.440	-93.447	344.351	158.737	-93.131	209.322	92.137	-93.432	204.837	92.435
106	Asp	-92.226	169.676	81.182	-92.456	169.689	84.734	-92.550	323.652	161.004	-92.585	249.891	120.200	-92.455	238.004	113.876
107	Phe	-93.200	116.048	61.156	-93.356	105.181	57.398	-93.135	207.928	114.812	-93.355	169.848	92.331	-93.004	158.087	83.413
110	Gln	-93.434	120.376	59.579	-93.599	111.230	57.194	-93.522	219.218	114.696	-93.595	178.158	89.972	-93.516	163.928	81.641
111	Ile	-93.890	95.079	56.075	-93.733	90.193	56.227	-93.651	178.761	108.831	-94.325	145.816	93.605	-94.267	133.606	83.807
112	Leu	-93.842	115.211	57.785	-93.729	108.171	58.697	-93.824	219.218	109.756	-94.058	174.066	84.987	-93.843	163.153	77.515
113	Arg	-93.650	150.673	74.997	-93.671	143.646	73.824	-93.654	279.432	142.324	-93.711	226.059	112.817	-93.399	211.781	102.650
115	Lys	-92.511	102.634	56.087	-93.175	96.706	53.469	-92.650	190.992	102.661	-92.694	146.045	79.107	-92.825	137.291	75.494
119	Asn	-92.378	214.798	112.812	-92.370	204.062	111.381	-92.392	396.097	218.425	-92.324	290.760	155.560	-92.375	276.071	146.675
120	Val	-93.600	191.991	102.780	-93.771	182.430	102.322	-93.685	352.818	194.361	-93.792	211.697	116.359	-93.744	204.496	109.274
122	Phe	-91.967	115.138	57.306	-92.123	108.950	58.338	-92.091	216.395	114.187	-91.967	175.560	102.255	-92.289	162.745	93.801
123	Leu	-93.733	130.424	67.544	-93.465	125.535	67.565	-93.704	248.384	134.624	-93.807	204.934	108.638	-93.701	192.099	99.760
124	Leu	-93.554	129.025	62.066	-93.374	124.859	65.313	-93.487	247.443	124.346	-92.985	198.587	90.689	-93.464	183.504	84.807
125	Val	-93.009	112.257	51.581	-93.002	105.498	54.410	-93.095	214.513	106.062	-93.048	175.009	85.656	-93.024	166.183	80.403
126	Gly	-93.878	93.332	49.978	-93.613	92.200	49.612	-93.673	177.820	95.925	-93.880	142.578	72.742	-93.211	140.192	71.223

Table C.1: $^{1}J_{NH}$ of RalAΔC-GDP from IPAP-HSQC and SOFAST-IPAP-HSQC

#	Resi	IPAP-HSQC									SOFAST-IPAP-HSQC					
		ns = 32 [A]			ns = 32 [B]			ns = 112			ns = 64 [A]			ns = 64 [B]		
		$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)
127	Asn	-94.211	129.246	60.929	-93.548	123.758	61.731	-94.037	241.798	117.366	-94.166	200.587	95.356	-93.967	187.904	86.505
128	Lys				-92.329	72.568	35.973				-92.343	103.343	55.347	-91.796	100.821	51.854
129	Ser	-93.041	102.500	52.514	-93.155	99.910	53.826	-93.280	190.992	100.828	-93.014	155.910	80.902	-93.442	152.952	77.633
130	Asp	-92.851	119.753	54.369	-92.985	113.824	54.150	-93.191	224.863	104.683	-93.414	178.311	83.341	-93.614	170.224	78.115
131	Leu	-91.962	129.703	63.951	-92.366	126.748	65.185	-91.948	246.502	128.714	-91.839	205.514	104.399	-92.021	190.320	96.567
132	Glu	-92.339	169.064	86.250	-92.202	171.389	91.998	-92.403	321.770	170.792	-92.662	163.538	83.366	-92.147	164.357	83.089
133	Asp	-93.572	163.291	88.415	-93.822	155.573	85.491	-93.648	303.894	167.448	-93.548	229.262	123.174	-93.667	221.982	118.465
134	Lys	-91.987	146.275	79.769	-91.664	152.948	84.205	-91.749	289.781	159.248	-91.706	234.449	121.115	-91.706	224.100	115.926
135	Arg	-94.400	141.978	69.850	-94.167	139.461	69.576	-94.474	269.083	135.657	-94.259	211.664	103.132	-94.270	202.236	97.550
136	Gln	-91.373	30.738	19.657	-90.258	28.417	16.011							-90.761	46.976	27.646
137	Val	-92.523	89.723	43.992	-92.356	89.888	44.997	-92.659	169.353	84.902	-92.716	133.219	67.091	-92.415	125.552	62.130
138	Ser	-94.596	113.376	58.034	-94.751	108.960	57.194	-94.485	215.454	110.977	-94.449	169.439	87.400	-94.602	165.751	85.027
140	Glu	-92.450	207.014	101.775	-92.465	204.158	102.020	-92.457	394.216	199.251	-92.538	298.063	146.957	-92.449	285.720	138.012
142	Ala	-94.286	160.270	68.918	-94.095	160.082	71.757	-94.173	305.776	138.274	-94.174	228.600	100.814	-94.177	219.816	98.230
143	Lys	-93.804	200.288	93.268	-93.609	188.979	91.650	-93.860	368.813	178.558	-93.749	281.071	130.311	-93.727	265.046	121.980
145	Arg	-93.463	157.105	70.308	-93.265	159.455	75.390	-93.494	301.072	141.152	-93.593	213.400	100.852	-93.326	207.256	93.107
146	Ala	-94.437	158.254	74.407	-94.371	158.526	75.572	-94.092	302.012	144.576	-94.573	240.479	119.584	-94.376	228.349	113.735
147	Asp	-93.693	197.166	93.524	-93.562	195.827	93.973	-93.619	377.280	184.044	-93.624	292.060	136.729	-93.664	282.637	132.834
148	Gln	-93.802	178.375	89.769	-93.551	178.215	92.483	-93.520	341.528	179.110	-93.527	269.035	140.816	-93.463	254.094	128.001
149	Trp	-92.854	124.555	65.180	-92.726	122.507	65.912	-93.057	238.035	123.581	-92.803	193.105	102.036	-93.046	185.353	94.939
151	Val	-92.775	193.797	97.561	-92.736	186.954	95.015	-92.801	358.463	180.107	-92.638	287.733	141.027	-92.835	277.319	134.695
152	Asn	-93.121	146.390	75.672	-93.191	141.904	77.087	-93.330	278.491	146.647	-93.329	218.081	111.476	-93.210	212.515	104.492
154	Val	-95.115	161.035	81.577	-95.138	158.715	81.701	-95.051	308.598	160.041	-95.180	249.200	129.915	-95.137	236.671	118.422
156	Thr	-94.169	110.756	56.786	-94.367	106.025	58.101	-94.257	206.046	108.248	-94.221	171.649	90.218	-94.358	163.472	85.075

Table C.1: $^{1}J_{NH}$ of RalAΔC-GDP from IPAP-HSQC and SOFAST-IPAP-HSQC

#	Resi	IPAP-HSQC									SOFAST-IPAP-HSQC					
		ns = 32 [A]			ns = 32 [B]			ns = 112			ns = 64 [A]			ns = 64 [B]		
		$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)
157	Ser	-93.079	116.664	58.220	-93.248	113.156	57.817	-93.085	218.277	112.305	-92.687	174.651	88.988	-93.059	167.200	83.786
158	Ala	-95.170	112.713	52.478	-95.290	107.227	54.565	-95.431	216.395	106.133	-95.654	178.162	84.093	-95.205	170.086	78.905
159	Lys	-91.849	98.435	56.194	-92.041	95.855	50.358	-92.392	183.465	97.972	-92.592	106.753	51.289	-91.796	105.844	50.968
160	Thr	-92.780	88.592	47.570	-92.421	88.346	47.394	-92.816	163.708	89.421	-92.798	137.401	70.337	-92.272	131.213	68.240
161	Arg	-92.490	97.531	46.587	-92.745	94.535	46.745	-92.677	185.347	91.387	-92.842	148.755	71.496	-92.660	139.616	66.897
162	Ala	-94.438	184.728	88.614	-94.370	184.731	91.018	-94.256	356.582	175.995	-94.192	273.320	131.639	-94.127	257.051	121.623
163	Asn	-93.726	117.923	63.186	-93.583	113.058	63.676	-93.909	222.040	122.893	-93.580	173.204	95.978	-93.662	168.786	90.931
164	Val	-93.295	146.514	67.682	-93.119	144.645	68.289	-93.097	279.432	133.443	-93.354	215.530	105.890	-93.159	205.453	98.178
167	Val	-94.149	121.180	59.895	-94.207	118.791	60.012	-93.814	221.099	112.069	-93.761	177.702	86.874	-93.740	172.493	85.117
168	Phe	-93.776	85.468	43.559	-93.556	81.631	42.004	-93.399	155.240	79.643	-93.165	73.788	38.825	-92.564	72.195	37.706
170	Asp	-94.082	145.323	70.800	-93.986	140.246	68.263	-93.961	273.787	137.900	-93.995	234.193	115.457	-94.203	223.468	104.854
171	Leu	-94.409	121.964	58.484	-94.841	116.021	55.068	-94.633	222.981	108.116	-94.279	188.998	92.693	-94.695	177.340	83.529
174	Glu	-91.060	162.413	97.289				-91.911	327.415	190.212						
176	Arg	-93.252	120.359	59.423	-93.122	115.492	58.637	-93.481	230.508	117.079	-93.357	176.854	88.120	-93.642	170.128	83.588
180	Met	-93.556	139.835	71.988	-93.822	107.591	60.059	-93.515	228.626	126.765	-94.499	184.832	101.041	-95.426	157.856	86.462
183	Ser	-93.023	260.341	204.672	-93.051	213.522	173.088	-92.992	438.435	352.916	-93.039	321.144	253.435	-93.026	286.328	224.283
184	Lys	-92.601	745.686	616.414	-92.626	686.103	583.143	-92.610	1373.639	1167.415	-92.629	1012.839	831.987	-92.627	928.013	748.345

Table C.2: $^1J_{NH}$ of RalAΔC-GDP from $^1J_{NH}$-modulated HSQC experiments
Conditions: 700 µM ^{15}N-RalAΔC-GDP, 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM $MgCl_2$, 0.05% NaN_3, 10% D_2O, 298 K.
C is the decay rate of the ^{15}N magnetisation as a result of transverse relaxation. $J_{2.5\%}$ and $J_{97.5\%}$ denote the 2.5% and 97.5% confidence intervals for the fitted J value, respectively. All units are in Hz.

Table C.2: $^1J_{NH}$ of RalAΔC-GDP from $^1J_{NH}$-modulated HSQC experiments

#	Resi	ns = 8, Δ = 11.60–37.20 ms				ns = 16*, Δ = 11.60–37.20 ms				ns = 32, Δ = 11.60–37.20 ms				ns = 32, Δ = 32.88–58.48 ms				ns = 8*, Δ = 32.88–58.48 ms			
		$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C
12	Leu	-92.894	-92.326	-93.496	29.247	-92.399	-92.060	-92.752	25.145	-92.715	-92.576	-92.858	26.603	-93.039	-92.861	-93.224	27.282	-92.942	-92.725	-93.170	26.039
13	Ala	-92.471	-92.155	-92.800	22.018	-92.750	-92.428	-93.086	21.986	-92.672	-92.475	-92.874	21.841	-92.882	-92.686	-93.088	21.548	-92.740	-92.569	-92.919	20.997
14	Leu	-92.120	-91.712	-92.552	17.231	-92.125	-91.749	-92.520	18.898	-92.178	-91.953	-92.409	17.910	-92.295	-92.065	-92.538	18.083	-92.045	-91.710	-92.411	14.002
15	His	-93.849	-93.218	-94.524	21.012	-94.047	-93.624	-94.490	21.782	-94.191	-94.017	-94.367	23.421	-93.982	-93.774	-94.199	22.383	-93.881	-93.681	-94.089	24.146
16	Lys	-91.562	-91.231	-91.907	21.646	-91.756	-91.377	-92.156	25.232	-91.672	-91.318	-92.043	21.702	-91.895	-91.741	-92.056	23.187	-91.805	-91.565	-92.060	24.477
17	Val	-93.334	-92.600	-94.131	25.868	-93.387	-92.973	-93.820	24.276	-93.474	-93.137	-93.823	25.489	-93.469	-93.292	-93.654	25.115	-93.339	-93.146	-93.540	26.480
18	Ile	-93.305	-92.868	-93.765	20.124	-93.392	-93.067	-93.730	22.764	-93.209	-92.926	-93.500	22.347	-93.436	-93.264	-93.615	21.104	-93.292	-93.142	-93.446	21.819
19	Met	-93.348	-92.527	-94.251	24.130	-93.049	-92.571	-93.555	25.361	-93.348	-93.142	-93.558	23.918	-93.272	-93.040	-93.516	24.353	-93.428	-92.937	-93.972	25.685
21	Gly	-93.194	-92.850	-93.552	19.831	-92.836	-92.434	-93.259	18.364	-93.035	-92.780	-93.298	19.480	-93.172	-93.062	-93.285	19.144	-93.098	-92.843	-93.367	19.705
22	Ser	-93.679	-93.271	-94.105	21.178	-93.712	-93.490	-93.939	20.684	-93.639	-93.438	-93.846	22.200	-93.923	-93.764	-94.088	21.625	-93.953	-93.765	-94.148	22.330
23	Gly	-94.288	-94.049	-94.533	19.306	-94.496	-94.304	-94.692	19.467	-94.372	-94.202	-94.545	19.469	-94.343	-94.193	-94.497	19.476	-94.642	-94.481	-94.807	19.368
24	Gly	-92.774	-92.164	-93.432	20.646	-92.686	-92.319	-93.070	20.419	-92.743	-92.425	-93.073	19.061	-92.901	-92.682	-93.130	19.579	-93.053	-92.889	-93.223	20.596
25	Val									-92.078	-90.354	-94.126	18.218								
26	Gly	-93.032	-92.593	-93.495	20.667	-93.078	-92.780	-93.386	21.327	-93.145	-92.949	-93.346	19.890	-93.200	-93.016	-93.392	20.805	-93.219	-93.012	-93.436	20.537
27	Lys	-90.941	-90.290	-91.649	21.939	-90.405	-89.957	-90.883	24.535	-90.140	-89.743	-90.560	22.372	-90.441	-90.192	-90.709	23.696	-90.604	-90.336	-90.893	24.845
28	Ser	-94.265	-93.994	-94.544	21.840	-93.848	-93.544	-94.162	22.583	-94.040	-93.795	-94.291	21.589	-94.160	-94.004	-94.320	23.069	-94.206	-94.039	-94.379	21.920
29	Ala	-93.808	-93.444	-94.186	20.647	-93.690	-93.420	-93.968	20.117	-93.501	-93.248	-93.760	20.349	-93.705	-93.562	-93.852	22.057	-93.806	-93.587	-94.035	21.357
30	Leu	-94.200	-93.955	-94.451	20.999	-94.065	-93.815	-94.321	21.242	-94.184	-93.992	-94.379	20.327	-94.250	-94.096	-94.409	23.254	-94.222	-94.085	-94.361	23.493
31	Thr	-92.793	-92.503	-93.094	25.735	-92.884	-92.700	-93.072	23.205	-92.804	-92.695	-92.915	24.083	-92.949	-92.825	-93.078	24.863	-92.742	-92.612	-92.877	23.420
32	Leu	-93.079	-92.508	-93.690	23.410	-92.900	-92.557	-93.258	21.277	-92.874	-92.601	-93.155	23.257	-93.092	-92.915	-93.276	25.463	-92.980	-92.726	-93.250	25.622

Table C.2: $^1J_{NH}$ of RalAΔC-GDP from $^1J_{NH}$-modulated HSQC experiments

#	Resi	ns = 8, Δ = 11.60–37.20 ms				ns = 16*, Δ = 11.60–37.20 ms				ns = 32, Δ = 11.60–37.20 ms				ns = 32, Δ = 32.88–58.48 ms				ns = 8*, Δ = 32.88–58.48 ms			
		$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C
33	Gln	-93.000	-92.574	-93.447	20.434	-93.022	-92.796	-93.253	19.448	-93.115	-92.910	-93.325	20.745	-93.195	-93.083	-93.311	20.902	-93.273	-93.067	-93.489	21.984
34	Phe	-93.495	-93.256	-93.741	19.025	-93.517	-93.408	-93.627	20.332	-93.496	-93.348	-93.646	19.605	-93.567	-93.454	-93.683	20.269	-93.577	-93.454	-93.703	21.419
35	Met	-92.730	-92.357	-93.121	21.257	-92.993	-92.755	-93.236	22.190	-92.839	-92.633	-93.051	22.156	-93.031	-92.848	-93.223	22.278	-92.997	-92.740	-93.271	21.223
36	Tyr					-90.682	-90.014	-91.387	18.650					-90.412	-90.247	-90.581	23.098				
38	Glu	-91.776	-91.496	-92.066	17.153	-91.435	-91.209	-91.669	17.476	-91.499	-91.289	-91.715	17.089	-91.811	-91.686	-91.939	19.302	-91.779	-91.705	-91.854	18.566
39	Phe	-92.212	-92.037	-92.392	20.281	-92.227	-92.041	-92.419	19.488	-92.186	-92.022	-92.354	18.595	-92.365	-92.208	-92.529	18.451	-92.215	-92.082	-92.352	18.600
40	Val	-94.577	-92.867	-96.547	24.954	-92.781	-91.724	-93.971	29.605	-93.071	-92.177	-94.048	35.799	-92.692	-92.047	-93.429	25.585	-92.746	-91.744	-93.995	23.367
41	Glu	-92.576	-92.237	-92.930	18.877	-92.662	-92.556	-92.769	19.300	-92.462	-92.262	-92.668	19.496	-92.682	-92.571	-92.797	19.288	-92.698	-92.549	-92.854	18.985
43	Tyr	-94.564	-93.825	-95.349	19.821	-94.094	-93.609	-94.605	22.951	-94.270	-93.928	-94.623	22.449	-94.095	-93.792	-94.416	22.528	-93.846	-93.520	-94.193	25.352
44	Glu	-93.547	-93.177	-93.933	22.535	-93.446	-93.243	-93.655	24.110	-93.232	-93.043	-93.425	23.621	-93.221	-93.075	-93.372	23.855	-93.298	-93.081	-93.525	23.888
46	Thr									-91.305	-90.214	-92.516	44.519								
47	Lys	-93.028	-92.305	-93.807	27.481	-92.073	-91.709	-92.454	19.776	-92.329	-91.946	-92.731	21.361	-92.590	-92.311	-92.889	22.738	-92.540	-92.238	-92.863	22.744
48	Ala	-91.838	-89.704	-94.542	42.682	-92.782	-92.249	-93.346	39.763	-92.831	-92.283	-93.415	39.683	-92.714	-91.717	-93.956	40.343	-93.069	-92.393	-93.831	42.944
49	Asp	-93.037	-92.989	-93.085	10.283	-93.011	-92.912	-93.112	10.186					-93.042	-92.956	-93.129	8.820	-93.036	-92.982	-93.090	9.176
50	Ser	-91.440	-91.106	-91.790	16.427	-91.597	-91.413	-91.786	18.572	-91.627	-91.456	-91.802	16.560	-91.632	-91.520	-91.748	17.182	-91.788	-91.678	-91.901	18.511
51	Tyr	-92.029	-91.565	-92.520	18.960	-92.368	-92.044	-92.706	20.690	-92.402	-92.160	-92.651	23.026	-92.400	-92.206	-92.603	23.268	-92.499	-92.226	-92.790	24.369
52	Arg	-92.682	-92.472	-92.898	17.428	-92.795	-92.555	-93.043	16.499	-92.791	-92.627	-92.959	18.233	-92.960	-92.893	-93.028	17.622	-92.959	-92.876	-93.044	18.258
53	Lys	-93.603	-93.292	-93.926	19.864	-93.529	-93.345	-93.716	20.189	-93.677	-93.520	-93.837	19.832	-93.801	-93.645	-93.962	20.055	-93.989	-93.853	-94.129	22.274
54	Lys	-92.460	-92.314	-92.609	17.595	-92.543	-92.360	-92.731	18.137	-92.404	-92.244	-92.568	18.567	-92.642	-92.560	-92.726	18.600	-92.605	-92.483	-92.732	18.647
56	Val	-92.018	-91.855	-92.183	16.783	-92.001	-91.856	-92.148	17.465	-92.009	-91.914	-92.106	17.940	-92.144	-92.075	-92.215	16.891	-92.127	-92.059	-92.195	16.902
57	Leu	-92.827	-92.512	-93.155	20.107	-92.771	-92.510	-93.040	21.893	-92.913	-92.733	-93.097	21.053	-92.949	-92.849	-93.051	21.929	-92.990	-92.736	-93.258	21.345
58	Asp	-94.822	-93.359	-96.444	29.153	-93.894	-92.765	-95.151	32.249	-94.336	-93.824	-94.873	35.449	-94.407	-93.683	-95.229	33.809	-93.502	-92.707	-94.416	36.200
59	Gly	-92.094	-91.900	-92.293	22.651	-91.925	-91.759	-92.095	23.147	-92.182	-91.985	-92.384	22.019	-92.147	-92.039	-92.258	20.886	-92.142	-91.990	-92.300	22.453
61	Glu	-92.816	-92.625	-93.011	18.490	-92.964	-92.764	-93.169	19.241	-92.977	-92.841	-93.115	19.465	-92.940	-92.854	-93.028	21.110	-92.968	-92.881	-93.057	19.861
62	Val	-93.327	-93.084	-93.576	19.040	-93.357	-92.952	-93.781	20.614	-93.446	-93.310	-93.585	21.183	-93.506	-93.276	-93.747	21.672	-93.351	-93.190	-93.519	19.580
63	Gln																	-92.665	-92.467	-92.873	19.928

Table C.2: $^1J_{NH}$ of RalAΔC-GDP from $^1J_{NH}$-modulated HSQC experiments

#	Resi	ns = 8, Δ = 11.60–37.20 ms				ns = 16*, Δ = 11.60–37.20 ms				ns = 32, Δ = 11.60–37.20 ms				ns = 32, Δ = 32.88–58.48 ms				ns = 8*, Δ = 32.88–58.48 ms			
		$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C
64	Ile	-93.094	-92.364	-93.882	22.642	-92.510	-92.163	-92.874	22.727	-92.677	-92.446	-92.914	24.089	-92.480	-92.355	-92.609	22.968	-92.548	-92.310	-92.800	23.232
65	Asp	-92.550	-92.032	-93.099	20.840	-92.806	-92.532	-93.090	19.627	-92.865	-92.623	-93.114	21.905	-92.839	-92.676	-93.008	21.711	-92.727	-92.400	-93.080	21.732
66	Ile	-92.654	-92.194	-93.141	21.763	-92.947	-92.628	-93.278	22.332	-93.019	-92.797	-93.247	21.840	-93.295	-93.132	-93.464	20.694	-93.391	-93.127	-93.671	22.018
67	Leu	-93.090	-92.697	-93.502	23.039	-93.409	-93.127	-93.700	21.665	-93.335	-93.147	-93.528	21.593	-93.455	-93.264	-93.653	20.436	-93.460	-93.311	-93.615	22.890
68	Asp	-93.355	-92.616	-94.160	21.983	-93.392	-93.119	-93.674	25.188	-93.397	-93.077	-93.728	24.545	-93.334	-93.163	-93.512	26.679	-93.588	-93.261	-93.938	27.220
69	Thr	-91.776	-90.368	-93.428	24.698	-92.218	-91.588	-92.903	25.855	-91.890	-91.219	-92.619	25.972	-92.372	-91.663	-93.212	28.322	-91.860	-91.026	-92.909	26.711
71	Gly	-93.695	-93.205	-94.211	23.609	-93.392	-93.135	-93.657	22.832	-93.431	-93.227	-93.639	23.490	-93.635	-93.459	-93.818	22.792	-93.584	-93.312	-93.871	24.294
72	Gln	-92.310	-91.487	-93.218	26.866	-92.188	-91.909	-92.477	24.702	-92.068	-91.766	-92.381	23.617	-92.051	-91.689	-92.448	27.218	-92.364	-92.012	-92.746	23.855
73	Glu	-92.543	-91.668	-93.508	24.847	-92.312	-92.075	-92.556	27.542	-92.344	-92.029	-92.671	27.294	-92.529	-92.321	-92.748	30.688	-92.275	-91.864	-92.730	26.940
74	Asp	-92.722	-92.502	-92.949	14.945	-92.600	-92.447	-92.756	15.814	-92.669	-92.527	-92.813	16.610	-92.654	-92.592	-92.718	14.547	-92.722	-92.631	-92.815	15.268
78	Ile	-91.468	-91.004	-91.961	20.334	-91.706	-91.428	-91.996	18.521	-91.821	-91.628	-92.019	20.498	-91.787	-91.493	-92.105	19.825	-91.846	-91.703	-91.995	22.460
79	Arg	-92.629	-91.798	-93.537	29.748	-92.749	-92.307	-93.215	27.258	-92.687	-92.326	-93.064	28.713	-92.649	-92.454	-92.853	30.426	-92.401	-91.985	-92.860	31.450
80	Asp	-93.368	-92.999	-93.752	22.767	-93.385	-93.195	-93.580	20.661	-93.267	-93.077	-93.461	21.467	-93.207	-93.069	-93.349	21.826	-93.321	-93.118	-93.533	22.076
83	Phe	-93.863	-93.418	-94.329	29.866	-93.900	-93.504	-94.314	27.065	-93.740	-93.517	-93.967	28.252	-93.947	-93.686	-94.221	30.037	-93.980	-93.712	-94.262	28.877
84	Arg	-93.104	-92.830	-93.387	21.865	-93.159	-92.922	-93.402	21.913	-93.289	-93.077	-93.505	22.178	-93.325	-93.203	-93.450	22.233	-93.267	-93.152	-93.386	23.365
85	Ser	-92.276	-91.677	-92.919	26.023	-92.008	-91.327	-92.754	29.119	-92.282	-91.944	-92.636	26.659	-92.343	-91.935	-92.795	28.234	-92.207	-91.525	-93.009	32.395
86	Gly	-95.325	-94.949	-95.714	21.866	-95.367	-95.091	-95.650	19.141	-95.321	-95.174	-95.470	19.545	-95.440	-95.294	-95.589	19.618	-95.454	-95.271	-95.642	20.628
88	Gly	-93.996	-93.704	-94.297	20.154	-93.889	-93.624	-94.161	19.573	-93.806	-93.604	-94.011	19.931	-93.796	-93.672	-93.923	20.219	-93.772	-93.595	-93.955	19.688
91	Cys	-93.861	-93.554	-94.177	21.132	-93.777	-93.444	-94.123	21.360	-93.860	-93.729	-93.992	20.917	-93.785	-93.706	-93.866	20.756	-93.759	-93.605	-93.918	21.424
92	Val	-94.114	-93.684	-94.562	19.172	-93.818	-93.491	-94.158	20.545	-93.745	-93.560	-93.933	20.015	-93.855	-93.702	-94.014	20.435	-93.933	-93.762	-94.111	21.101
93	Phe	-93.994	-93.730	-94.264	21.392	-94.295	-94.075	-94.519	20.550	-94.255	-94.037	-94.478	21.838	-94.429	-94.322	-94.538	22.581	-94.295	-94.134	-94.461	21.396
94	Ser	-91.982	-91.379	-92.632	23.244	-91.489	-91.280	-91.703	22.802	-91.505	-91.214	-91.810	22.759	-92.024	-91.814	-92.245	24.461	-92.009	-91.666	-92.384	21.048
95	Ile	-94.802	-94.379	-95.242	23.741	-94.982	-94.735	-95.235	24.496	-94.669	-94.495	-94.845	23.009	-94.821	-94.614	-95.036	24.636	-95.092	-94.871	-95.321	26.088
96	Thr	-92.830	-92.493	-93.180	21.314	-92.754	-92.515	-93.000	20.801	-92.767	-92.619	-92.919	21.365	-92.907	-92.803	-93.013	22.714	-92.812	-92.662	-92.968	22.811
98	Met	-93.406	-92.970	-93.863	26.433	-93.171	-92.704	-93.663	28.107	-93.238	-92.888	-93.603	26.463	-93.272	-93.016	-93.542	27.155	-93.527	-93.080	-94.016	29.790
99	Glu	-92.141	-91.640	-92.673	28.824	-92.210	-91.856	-92.579	29.143	-92.165	-91.838	-92.507	28.420	-92.416	-92.183	-92.663	26.371	-92.537	-92.200	-92.902	32.354

Table C.2: $^1J_{NH}$ of RalAΔC-GDP from $^1J_{NH}$-modulated HSQC experiments

#	Resi	ns = 8, Δ = 11.60–37.20 ms				ns = 16*, Δ = 11.60–37.20 ms				ns = 32, Δ = 11.60–37.20 ms				ns = 32, Δ = 32.88–58.48 ms				ns = 8*, Δ = 32.88–58.48 ms			
		$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C
100	Ser	-91.487	-91.054	-91.946	22.678	-91.269	-91.017	-91.531	21.532	-91.227	-91.103	-91.353	20.903	-91.417	-91.286	-91.553	23.003	-91.320	-91.179	-91.466	22.748
101	Phe	-92.365	-92.009	-92.739	19.814	-92.391	-92.117	-92.675	22.572	-92.329	-92.172	-92.490	21.586	-92.536	-92.391	-92.686	23.276	-92.550	-92.411	-92.693	22.132
102	Ala	-93.763	-93.429	-94.110	20.766	-93.659	-93.518	-93.802	19.884	-93.670	-93.483	-93.861	20.638	-93.692	-93.584	-93.802	20.784	-93.649	-93.486	-93.818	21.820
103	Ala	-92.801	-92.558	-93.051	18.159	-92.635	-92.495	-92.777	18.725	-92.608	-92.517	-92.700	18.534	-92.736	-92.663	-92.809	19.458	-92.696	-92.609	-92.784	20.028
104	Thr	-92.704	-92.351	-93.073	19.043	-92.726	-92.459	-93.003	20.086	-92.730	-92.559	-92.904	20.375	-93.016	-92.892	-93.142	21.027	-92.917	-92.790	-93.047	20.907
105	Ala	-92.819	-92.499	-93.153	20.356	-92.955	-92.764	-93.151	18.667	-92.887	-92.681	-93.099	18.596	-93.093	-92.999	-93.190	20.431	-93.019	-92.919	-93.122	17.851
106	Asp	-92.312	-92.110	-92.520	17.648	-92.153	-91.949	-92.362	17.677	-92.151	-92.006	-92.298	18.086	-92.246	-92.144	-92.352	18.610	-92.290	-92.136	-92.451	18.191
109	Glu					-92.998	-92.809	-93.191	19.892												
110	Gln	-93.067	-92.742	-93.404	19.625	-93.287	-93.104	-93.475	21.714	-93.209	-93.064	-93.356	21.380	-93.257	-93.138	-93.379	20.338	-93.268	-93.135	-93.406	21.583
112	Leu	-93.686	-93.265	-94.126	21.253	-93.584	-93.258	-93.921	21.874	-93.831	-93.582	-94.086	22.906	-93.659	-93.532	-93.790	17.533	-93.551	-93.417	-93.689	22.268
113	Arg	-93.235	-92.974	-93.504	20.101	-93.375	-93.226	-93.528	19.956	-93.382	-93.166	-93.603	20.994	-93.406	-93.316	-93.498	21.996	-93.469	-93.360	-93.579	21.932
114	Val	-92.614	-92.362	-92.875	21.137	-92.578	-92.399	-92.762	20.501	-92.750	-92.660	-92.841	20.069	-92.838	-92.702	-92.978	21.702	-92.903	-92.776	-93.032	20.330
115	Lys	-92.387	-91.913	-92.890	22.724	-92.417	-92.209	-92.631	21.693	-92.521	-92.326	-92.721	24.195	-92.717	-92.609	-92.828	26.479	-92.659	-92.461	-92.866	25.387
116	Glu					-92.792	-92.608	-92.979	23.218					-92.832	-92.684	-92.983	23.133	-92.882	-92.720	-93.051	24.734
117	Asp	-90.725	-90.391	-91.076	13.451	-90.857	-90.433	-91.310	14.632	-90.955	-90.678	-91.242	14.972	-91.112	-90.975	-91.255	15.970	-91.088	-90.929	-91.254	14.856
119	Asn	-92.059	-91.864	-92.258	19.153	-91.998	-91.884	-92.115	19.577	-92.053	-91.929	-92.179	19.007	-92.092	-92.016	-92.169	18.900	-92.108	-92.015	-92.204	18.892
120	Val	-93.208	-92.909	-93.517	14.541	-93.227	-93.073	-93.385	14.488	-93.189	-93.028	-93.353	14.532	-93.386	-93.311	-93.462	15.807	-93.450	-93.360	-93.541	15.714
122	Phe	-91.607	-91.135	-92.108	21.024	-91.467	-91.192	-91.753	19.643	-91.358	-91.198	-91.521	20.855	-91.479	-91.368	-91.593	20.816	-91.618	-91.423	-91.825	20.082
124	Leu	-92.586	-92.217	-92.973	18.747	-92.775	-92.519	-93.040	21.551	-92.581	-92.397	-92.769	21.212	-92.812	-92.638	-92.993	22.390	-92.770	-92.634	-92.910	21.418
125	Val	-92.548	-92.200	-92.911	20.525	-92.412	-92.267	-92.560	20.839	-92.545	-92.343	-92.752	21.243	-92.544	-92.468	-92.622	21.118	-92.683	-92.523	-92.849	21.736
126	Gly	-93.614	-93.031	-94.235	23.199	-93.464	-93.213	-93.722	21.841	-93.664	-93.359	-93.979	21.811	-93.552	-93.436	-93.671	21.604	-93.449	-93.203	-93.708	21.786
127	Asn	-93.632	-93.378	-93.893	22.311	-93.806	-93.585	-94.031	20.568	-93.864	-93.662	-94.069	22.615	-93.808	-93.653	-93.968	21.968	-93.762	-93.638	-93.890	19.451
128	Lys	-91.675	-90.793	-92.665	23.270	-91.811	-91.251	-92.417	22.922	-91.603	-91.154	-92.083	22.827	-91.611	-91.242	-92.020	22.394	-91.816	-91.305	-92.398	24.036
129	Ser	-92.775	-92.381	-93.188	22.514	-92.799	-92.572	-93.032	22.501	-92.883	-92.650	-93.122	23.018	-92.690	-92.543	-92.842	23.109	-92.619	-92.503	-92.738	23.699
130	Asp	-92.563	-92.255	-92.884	23.218	-92.606	-92.382	-92.836	21.312	-92.631	-92.480	-92.784	21.750	-92.760	-92.602	-92.924	21.451	-92.652	-92.484	-92.827	23.045
131	Leu	-92.111	-91.790	-92.445	21.206	-91.832	-91.711	-91.956	19.647	-91.853	-91.671	-92.040	21.833	-91.947	-91.841	-92.057	22.738	-91.858	-91.711	-92.011	22.623

Table C.2: $^1J_{NH}$ of RalAΔC-GDP from $^1J_{NH}$-modulated HSQC experiments

#	Resi	ns = 8, Δ = 11.60–37.20 ms				ns = 16*, Δ = 11.60–37.20 ms				ns = 32, Δ = 11.60–37.20 ms				ns = 32, Δ = 32.88–58.48 ms				ns = 8*, Δ = 32.88–58.48 ms			
		$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C
132	Glu	-91.521	-91.270	-91.782	15.759	-91.666	-91.513	-91.822	16.115	-91.634	-91.480	-91.792	16.088	-91.876	-91.805	-91.947	16.631	-91.880	-91.746	-92.020	16.717
133	Asp	-93.223	-93.034	-93.416	21.559	-93.395	-93.224	-93.570	23.038	-93.370	-93.228	-93.515	22.122	-93.364	-93.260	-93.470	23.609	-93.354	-93.215	-93.497	23.192
134	Lys	-91.468	-91.205	-91.742	18.361	-91.484	-91.332	-91.638	18.168	-91.502	-91.374	-91.632	19.144	-91.506	-91.403	-91.612	20.173	-91.567	-91.455	-91.683	19.680
135	Arg	-93.830	-93.548	-94.121	20.947	-93.837	-93.679	-93.998	19.486	-93.869	-93.687	-94.053	19.501	-94.135	-94.029	-94.244	19.896	-93.989	-93.829	-94.154	20.000
136	Gln	-89.540	-87.344	-92.267	20.559	-90.155	-89.611	-90.744	36.480	-90.042	-89.165	-91.034	32.341								
137	Val	-92.261	-91.869	-92.672	21.475	-92.034	-91.795	-92.280	20.096	-92.077	-91.857	-92.304	19.956	-92.369	-92.178	-92.570	20.786	-92.198	-92.039	-92.363	19.244
138	Ser	-93.965	-93.607	-94.337	22.717	-93.957	-93.627	-94.298	24.204	-94.105	-93.859	-94.357	23.705	-94.361	-94.155	-94.575	23.598	-94.387	-94.169	-94.613	24.720
140	Glu	-92.048	-91.902	-92.196	18.330	-92.263	-92.153	-92.375	19.907	-92.142	-92.021	-92.264	18.957	-92.255	-92.163	-92.349	20.476	-92.229	-92.131	-92.328	20.115
142	Ala	-93.691	-93.504	-93.881	19.059	-93.720	-93.577	-93.866	19.928	-93.702	-93.618	-93.786	19.630	-93.927	-93.830	-94.026	20.582	-93.972	-93.852	-94.095	19.882
143	Lys	-93.404	-93.287	-93.523	19.758	-93.464	-93.366	-93.563	19.794	-93.427	-93.284	-93.572	19.570	-93.519	-93.420	-93.619	20.493	-93.495	-93.387	-93.606	21.523
144	Asn	-93.218	-93.005	-93.436	18.772	-93.322	-93.178	-93.468	18.649	-93.372	-93.262	-93.483	18.676	-93.497	-93.386	-93.610	19.709	-93.607	-93.510	-93.706	20.024
145	Arg	-92.978	-92.657	-93.312	19.729	-92.878	-92.635	-93.129	19.139	-92.918	-92.793	-93.044	19.477	-93.109	-93.038	-93.182	21.140	-93.074	-92.916	-93.238	19.558
146	Ala	-93.319	-93.049	-93.597	21.086	-93.449	-93.252	-93.650	20.381	-93.285	-93.099	-93.474	20.549	-93.573	-93.476	-93.672	20.866	-93.548	-93.379	-93.723	19.922
147	Asp	-93.293	-93.101	-93.490	18.958	-93.395	-93.202	-93.593	19.378	-93.339	-93.186	-93.494	19.168	-93.361	-93.272	-93.453	19.612	-93.371	-93.266	-93.479	19.876
149	Trp	-92.859	-92.623	-93.102	21.100	-92.750	-92.632	-92.869	18.867	-92.871	-92.623	-93.127	19.515	-92.776	-92.681	-92.873	20.338	-92.942	-92.811	-93.076	20.073
151	Val	-92.511	-92.285	-92.743	19.331	-92.345	-92.144	-92.553	18.984	-92.463	-92.316	-92.613	18.557	-92.514	-92.448	-92.581	19.182	-92.582	-92.497	-92.668	19.022
152	Asn	-93.058	-92.813	-93.310	20.511	-92.923	-92.753	-93.095	19.230	-92.994	-92.839	-93.153	18.594	-93.005	-92.911	-93.101	21.444	-92.939	-92.825	-93.056	21.252
153	Tyr	-91.529	-91.168	-91.908	20.038	-91.386	-91.207	-91.569	18.449	-91.503	-91.347	-91.663	19.462	-91.478	-91.369	-91.590	19.179	-91.496	-91.316	-91.685	17.818
156	Thr	-93.927	-93.628	-94.236	22.735	-94.272	-93.914	-94.643	22.930	-94.168	-94.026	-94.312	22.563	-93.953	-93.801	-94.110	22.142	-94.109	-93.991	-94.230	22.087
157	Ser	-92.367	-91.834	-92.938	20.239	-92.687	-92.466	-92.915	21.568	-92.628	-92.451	-92.810	20.260	-92.599	-92.450	-92.754	21.579	-92.548	-92.463	-92.634	20.108
158	Ala	-94.876	-94.519	-95.244	22.091	-94.725	-94.412	-95.048	21.859	-94.816	-94.627	-95.008	20.745	-94.877	-94.686	-95.074	21.827	-95.067	-94.911	-95.227	22.134
159	Lys	-92.041	-91.711	-92.385	19.627	-92.168	-91.829	-92.521	20.524	-92.285	-92.092	-92.484	19.208	-92.455	-92.293	-92.623	21.695	-92.612	-92.318	-92.928	21.667
160	Thr	-92.464	-92.173	-92.765	21.456	-92.254	-91.990	-92.526	21.218	-92.335	-92.159	-92.516	20.718	-92.460	-92.328	-92.597	22.282	-92.398	-92.245	-92.557	21.429
161	Arg	-92.558	-92.224	-92.907	21.114	-92.531	-92.300	-92.768	20.879	-92.572	-92.435	-92.712	20.791	-92.445	-92.312	-92.583	20.873	-92.537	-92.344	-92.740	20.445
162	Ala	-94.094	-93.932	-94.258	17.646	-93.962	-93.781	-94.147	17.829	-94.098	-93.964	-94.234	18.227	-93.994	-93.898	-94.092	18.260	-93.994	-93.907	-94.083	18.544
163	Asn	-93.365	-93.102	-93.636	20.732	-93.498	-93.339	-93.660	21.594	-93.481	-93.288	-93.677	20.997	-93.479	-93.361	-93.600	20.686	-93.532	-93.385	-93.684	21.838

Table C.2: $^1J_{NH}$ of RalAΔC-GDP from $^1J_{NH}$-modulated HSQC experiments

#	Resi	ns = 8, Δ = 11.60–37.20 ms				ns = 16*, Δ = 11.60–37.20 ms				ns = 32, Δ = 11.60–37.20 ms				ns = 32, Δ = 32.88–58.48 ms				ns = 8*, Δ = 32.88–58.48 ms			
		$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$	$J_{2.5\%}$	$J_{97.5\%}$	C
164	Val	-92.645	-92.394	-92.903	18.476	-92.740	-92.554	-92.930	19.160	-92.714	-92.598	-92.831	19.653	-92.900	-92.789	-93.015	20.207	-92.873	-92.727	-93.024	20.322
165	Asp	-92.752	-92.505	-93.006	19.341	-92.692	-92.568	-92.818	17.625	-92.520	-92.406	-92.636	18.407	-92.746	-92.640	-92.855	18.910	-92.689	-92.570	-92.812	19.555
167	Val	-93.425	-93.077	-93.787	20.734	-93.562	-93.349	-93.779	20.310	-93.744	-93.639	-93.850	20.735	-93.752	-93.623	-93.883	22.404	-93.685	-93.514	-93.863	22.542
168	Phe	-93.031	-92.668	-93.411	22.159	-92.702	-92.403	-93.012	19.711	-92.788	-92.572	-93.010	21.266	-93.044	-92.875	-93.220	21.113	-93.019	-92.841	-93.204	21.871
170	Asp	-93.811	-93.529	-94.102	20.707	-93.733	-93.532	-93.939	20.610	-93.664	-93.513	-93.816	19.178	-93.858	-93.753	-93.965	21.581	-93.845	-93.768	-93.922	21.887
171	Leu	-94.136	-93.875	-94.405	20.379	-94.186	-93.941	-94.437	20.100	-94.245	-94.073	-94.421	20.594	-94.235	-94.118	-94.354	21.275	-94.288	-94.155	-94.425	21.903
172	Met	-93.947	-93.610	-94.297	22.207	-93.958	-93.717	-94.205	21.710	-93.925	-93.709	-94.147	21.797	-93.974	-93.904	-94.044	23.335	-93.947	-93.823	-94.075	23.806
176	Arg	-93.078	-92.803	-93.362	23.416	-93.069	-92.930	-93.210	23.169	-93.036	-92.769	-93.311	21.243	-93.185	-93.044	-93.331	23.035	-93.156	-93.053	-93.261	22.566
177	Ala	-93.123	-92.889	-93.364	20.161	-93.316	-93.161	-93.474	20.903	-93.175	-93.037	-93.315	20.135	-93.265	-93.197	-93.333	20.723	-93.244	-93.152	-93.337	21.844
179	Lys	-92.928	-92.602	-93.267	20.040	-93.010	-92.784	-93.243	21.786	-92.838	-92.640	-93.042	21.444	-92.973	-92.855	-93.094	22.252	-92.954	-92.792	-93.122	21.881
180	Met	-93.116	-92.854	-93.385	18.193	-93.045	-92.912	-93.179	18.710	-92.992	-92.842	-93.144	18.267	-93.176	-93.086	-93.267	18.686	-93.048	-92.941	-93.158	19.244
181	Glu	-92.788	-92.572	-93.010	15.136	-92.778	-92.634	-92.925	15.012	-92.831	-92.707	-92.956	14.874	-92.929	-92.888	-92.970	16.361	-92.940	-92.873	-93.008	16.138
182	Asp	-92.840	-92.746	-92.936	12.669																
183	Ser	-92.575	-92.452	-92.699	12.222	-92.640	-92.512	-92.771	12.139	-92.644	-92.535	-92.754	11.910	-92.724	-92.694	-92.754	12.549	-92.719	-92.660	-92.779	12.232
184	Lys	-92.306	-92.226	-92.388	3.158	-92.296	-92.223	-92.370	3.524	-92.306	-92.235	-92.377	3.465	-92.381	-92.357	-92.406	3.874	-92.373	-92.327	-92.420	3.960

Table C.3: $^{1}J_{NH}$ of RalAΔC-GDP from $^{1}J_{NH}$-modulated TROSY-HSQC experiments
Conditions: 700 µM ^{15}N-RalAΔC-GDP, 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM $MgCl_2$, 0.05% NaN_3, 10% D_2O, 298 K.
C is the decay rate of the ^{15}N magnetisation as a result of transverse relaxation.
$J_{2.5\%}$ and $J_{97.5\%}$ denote the 2.5% and 97.5% confidence intervals for the fitted J value, respectively.

Table C.3: $^{1}J_{NH}$ of RalAΔC-GDP from $^{1}J_{NH}$-modulated TROSY-HSQC experiments

#	Resi	ns = 64*, Δ = 32.88–58.48 ms				ns = 48*, Δ = 22.24–47.84 ms			
		$^{1}J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C (s^{-1})	$^{1}J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C (s^{-1})
12	Leu	-93.283	-92.719	-93.924	28.879	-92.982	-92.594	-93.394	30.278
13	Ala	-92.718	-92.352	-93.116	21.992	-92.906	-92.559	-93.273	22.749
14	Leu	-92.130	-91.959	-92.309	19.485	-92.278	-92.098	-92.465	18.285
15	His	-94.030	-93.758	-94.317	25.352	-94.017	-93.672	-94.381	23.475
16	Lys	-91.871	-91.579	-92.187	26.446	-91.757	-91.126	-92.471	23.878
17	Val	-93.543	-93.198	-93.913	25.373	-93.487	-93.187	-93.800	26.011
18	Ile	-93.334	-93.156	-93.520	20.631	-93.415	-93.210	-93.627	22.971
19	Met	-93.425	-93.193	-93.669	26.549	-93.430	-93.223	-93.644	23.987
21	Gly	-93.138	-93.017	-93.263	19.404	-93.186	-93.025	-93.352	18.697
22	Ser	-94.071	-93.937	-94.208	23.763	-94.038	-93.698	-94.396	22.417
23	Gly	-94.513	-94.389	-94.640	20.217	-94.411	-94.213	-94.613	20.392
24	Gly	-92.828	-92.669	-92.994	19.417	-92.873	-92.596	-93.163	20.208
26	Gly	-93.164	-93.002	-93.332	21.428	-93.298	-93.131	-93.470	21.865
27	Lys	-90.522	-89.916	-91.222	28.264	-90.775	-90.368	-91.218	26.488
28	Ser	-94.182	-93.981	-94.391	25.300	-94.224	-93.968	-94.488	22.807
29	Ala	-93.982	-93.775	-94.199	21.963	-93.692	-93.505	-93.884	20.465
30	Leu	-94.362	-94.180	-94.549	23.266	-94.454	-94.217	-94.699	22.708
31	Thr	-93.050	-92.901	-93.204	22.969	-93.033	-92.901	-93.168	24.055
32	Leu	-93.094	-92.789	-93.421	24.326	-93.326	-93.018	-93.649	25.510

Table C.3: $^1J_{NH}$ of RalAΔC-GDP from $^1J_{NH}$-modulated TROSY-HSQC experiments

#	Resi	ns = 64*, Δ = 32.88–58.48 ms				ns = 48*, Δ = 22.24–47.84 ms			
		$^1J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C (s^{-1})	$^1J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C (s^{-1})
33	Gln	-93.303	-93.137	-93.474	19.620	-93.337	-93.088	-93.596	19.939
34	Phe	-93.711	-93.554	-93.873	20.399	-93.683	-93.531	-93.839	19.816
35	Met	-93.013	-92.788	-93.250	25.806	-93.049	-92.773	-93.337	21.477
38	Glu	-91.856	-91.648	-92.075	17.002	-91.855	-91.668	-92.050	17.711
39	Phe	-92.366	-92.238	-92.499	20.246	-92.360	-92.222	-92.502	20.397
40	Val					-93.599	-91.374	-96.462	33.908
41	Glu	-92.836	-92.688	-92.989	19.355	-92.811	-92.661	-92.966	19.610
43	Tyr	-94.202	-93.802	-94.631	20.210	-94.225	-93.591	-94.917	23.488
44	Glu	-93.176	-93.016	-93.342	22.802	-93.218	-92.935	-93.515	22.781
47	Lys	-92.662	-92.436	-92.899	23.426	-92.547	-92.105	-93.026	24.877
48	Ala					-92.325	-91.414	-93.375	45.348
49	Asp	-93.206	-93.136	-93.278	9.873	-93.268	-93.149	-93.389	10.483
50	Ser	-91.714	-91.558	-91.876	18.154	-91.750	-91.631	-91.872	16.377
51	Tyr	-92.559	-92.320	-92.812	21.494	-92.455	-92.178	-92.745	22.104
52	Arg	-92.948	-92.815	-93.086	16.997	-92.937	-92.813	-93.064	18.954
53	Lys	-93.939	-93.806	-94.075	20.875	-93.901	-93.758	-94.048	21.467
54	Lys	-92.664	-92.554	-92.777	18.346	-92.573	-92.406	-92.744	19.455
56	Val	-92.149	-92.064	-92.236	16.200	-92.261	-92.137	-92.388	18.137
57	Leu	-92.997	-92.829	-93.172	20.890	-92.958	-92.781	-93.141	22.838
58	Asp	-94.268	-93.244	-95.416	29.519	-94.476	-93.509	-95.549	36.233
59	Gly	-92.171	-92.035	-92.312	23.535	-92.163	-92.061	-92.268	22.847
61	Glu	-93.011	-92.916	-93.109	18.953	-92.974	-92.842	-93.108	19.514
62	Val	-93.583	-93.370	-93.806	21.192	-93.672	-93.438	-93.914	20.714
64	Ile	-92.690	-92.410	-92.990	24.193	-92.561	-92.223	-92.921	24.637
65	Asp	-92.902	-92.696	-93.118	21.028	-93.263	-93.051	-93.483	23.627
66	Ile	-93.285	-93.050	-93.533	23.962	-93.285	-93.080	-93.497	22.611

Table C.3: $^{1}J_{NH}$ of RalAΔC-GDP from $^{1}J_{NH}$-modulated TROSY-HSQC experiments

#	Resi	ns = 64*, Δ = 32.88–58.48 ms				ns = 48*, Δ = 22.24–47.84 ms			
		$^{1}J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C (s^{-1})	$^{1}J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C (s^{-1})
67	Leu	-93.482	-93.238	-93.739	20.941	-93.553	-93.340	-93.773	22.730
68	Asp	-93.906	-93.671	-94.152	26.719	-93.459	-93.098	-93.840	27.038
69	Thr	-92.278	-91.343	-93.432	29.615	-91.370	-90.452	-92.476	26.713
71	Gly	-93.822	-93.550	-94.108	25.155	-93.521	-93.203	-93.854	24.726
72	Gln	-91.935	-91.473	-92.451	27.194	-92.249	-91.660	-92.903	29.585
73	Glu	-92.875	-92.207	-93.641	31.344	-92.307	-91.857	-92.794	29.640
74	Asp	-92.892	-92.728	-93.062	16.096	-92.906	-92.792	-93.022	17.053
78	Ile	-91.910	-91.698	-92.135	20.043	-92.012	-91.745	-92.291	23.537
79	Arg	-92.223	-91.666	-92.863	31.039	-92.871	-92.445	-93.327	32.431
80	Asp	-93.422	-93.241	-93.611	23.519	-93.535	-93.215	-93.870	24.005
83	Phe	-94.012	-93.482	-94.597	27.873	-93.945	-93.603	-94.303	28.799
84	Arg	-93.346	-93.184	-93.515	21.046				
85	Ser	-92.362	-91.705	-93.136	25.701	-92.654	-92.301	-93.028	31.184
86	Gly	-95.605	-95.421	-95.793	19.256	-95.554	-95.290	-95.826	20.142
88	Gly	-93.861	-93.632	-94.102	20.496	-93.994	-93.734	-94.263	20.324
91	Cys	-93.851	-93.720	-93.986	23.179	-93.786	-93.621	-93.954	21.043
92	Val	-93.930	-93.794	-94.070	22.539	-93.804	-93.604	-94.010	22.832
93	Phe	-94.420	-94.248	-94.596	21.293	-94.308	-94.120	-94.502	23.070
94	Ser	-91.823	-91.445	-92.242	22.012	-92.056	-91.612	-92.542	21.486
95	Ile	-94.825	-94.631	-95.026	22.483	-94.855	-94.595	-95.124	24.343
96	Thr	-92.859	-92.758	-92.962	24.720	-92.984	-92.819	-93.154	22.851
98	Met	-93.460	-93.187	-93.749	28.334	-93.188	-92.832	-93.565	27.566
99	Glu	-92.218	-91.838	-92.636	27.456	-92.394	-92.221	-92.573	29.600
100	Ser	-91.381	-91.272	-91.494	24.179	-91.291	-91.138	-91.449	22.495
101	Phe	-92.579	-92.288	-92.891	23.369	-92.435	-92.234	-92.644	23.514
102	Ala	-93.710	-93.492	-93.937	20.964	-93.789	-93.614	-93.968	20.819

Table C.3: $^1J_{NH}$ of RalAΔC-GDP from $^1J_{NH}$-modulated TROSY-HSQC experiments

#	Resi	ns = 64*, Δ = 32.88–58.48 ms				ns = 48*, Δ = 22.24–47.84 ms			
		$^1J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C (s^{-1})	$^1J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C (s^{-1})
103	Ala	-92.708	-92.612	-92.807	19.656	-92.710	-92.586	-92.837	18.932
104	Thr	-92.944	-92.821	-93.069	21.780	-93.007	-92.797	-93.225	22.153
105	Ala	-92.998	-92.875	-93.125	19.808	-93.068	-92.956	-93.182	20.037
106	Asp	-92.351	-92.223	-92.484	19.283	-92.377	-92.263	-92.494	19.706
110	Gln	-93.295	-93.105	-93.494	22.253	-93.147	-92.992	-93.306	22.704
112	Leu	-93.517	-93.376	-93.662	16.621	-93.633	-93.438	-93.835	20.943
113	Arg	-93.477	-93.359	-93.598	21.393	-93.380	-93.185	-93.581	21.540
114	Val	-92.777	-92.688	-92.868	21.974	-92.815	-92.576	-93.063	20.835
115	Lys	-92.752	-92.484	-93.037	23.733	-92.690	-92.364	-93.036	24.630
117	Asp	-91.201	-91.069	-91.339	16.330	-90.865	-90.670	-91.069	14.657
119	Asn	-92.180	-92.051	-92.313	19.559	-92.193	-92.097	-92.291	19.085
120	Val	-93.388	-93.282	-93.495	15.560	-93.356	-93.217	-93.498	15.763
122	Phe	-91.525	-91.325	-91.736	19.123	-91.510	-91.227	-91.810	21.276
124	Leu	-92.773	-92.650	-92.899	20.427	-92.899	-92.759	-93.042	22.210
125	Val	-92.692	-92.524	-92.867	20.408	-92.671	-92.517	-92.830	21.537
126	Gly	-93.610	-93.402	-93.828	22.353	-93.655	-93.472	-93.844	24.285
127	Asn	-93.990	-93.765	-94.224	21.875	-93.942	-93.780	-94.109	22.198
128	Lys	-91.671	-91.384	-91.981	24.270	-91.855	-91.387	-92.367	25.319
129	Ser	-92.820	-92.615	-93.036	23.556	-92.887	-92.708	-93.072	23.445
130	Asp	-93.033	-92.900	-93.170	22.211	-92.881	-92.740	-93.025	22.237
131	Leu	-92.081	-91.855	-92.322	22.634	-91.857	-91.718	-92.000	23.011
132	Glu	-91.899	-91.766	-92.038	16.354	-91.924	-91.739	-92.116	17.142
133	Asp	-93.423	-93.268	-93.583	22.803	-93.444	-93.340	-93.549	23.619
134	Lys	-91.539	-91.427	-91.655	19.251	-91.522	-91.393	-91.655	19.589
135	Arg	-94.049	-93.870	-94.234	19.850	-94.027	-93.833	-94.227	20.881
137	Val	-92.300	-92.045	-92.572	19.278	-92.390	-92.110	-92.686	21.077

Table C.3: $^1J_{NH}$ of RalAΔC-GDP from $^1J_{NH}$-modulated TROSY-HSQC experiments

#	Resi	ns = 64*, Δ = 32.88–58.48 ms				ns = 48*, Δ = 22.24–47.84 ms			
		$^1J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C (s^{-1})	$^1J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C (s^{-1})
138	Ser	-94.306	-94.153	-94.463	24.413	-94.359	-94.145	-94.580	23.943
140	Glu	-92.301	-92.172	-92.435	20.247	-92.328	-92.168	-92.492	19.988
142	Ala	-94.020	-93.858	-94.187	20.434	-94.011	-93.872	-94.153	20.911
143	Lys	-93.599	-93.490	-93.711	19.368	-93.536	-93.364	-93.713	21.216
144	Asn	-93.582	-93.481	-93.686	20.252	-93.565	-93.432	-93.702	19.616
145	Arg	-93.162	-93.050	-93.278	20.868	-93.198	-93.059	-93.341	21.089
146	Ala	-93.645	-93.457	-93.841	21.849	-93.513	-93.362	-93.667	19.974
147	Asp	-93.421	-93.308	-93.537	19.276	-93.420	-93.277	-93.567	20.769
149	Trp	-92.796	-92.654	-92.944	20.154	-92.958	-92.727	-93.198	21.139
151	Val	-92.542	-92.434	-92.653	19.956	-92.663	-92.524	-92.806	20.188
152	Asn	-93.103	-92.961	-93.249	20.827	-93.083	-92.958	-93.212	20.620
153	Tyr	-91.398	-91.194	-91.612	19.639	-91.527	-91.379	-91.679	18.575
156	Thr	-94.134	-94.024	-94.247	22.580	-94.273	-94.120	-94.428	21.912
157	Ser	-92.626	-92.509	-92.746	20.445	-92.635	-92.426	-92.851	20.442
158	Ala	-94.951	-94.730	-95.181	22.104	-95.078	-94.899	-95.261	21.748
159	Lys	-92.484	-92.269	-92.710	20.761	-92.462	-92.227	-92.708	20.743
160	Thr	-92.545	-92.345	-92.756	20.232	-92.601	-92.396	-92.815	21.834
161	Arg	-92.450	-92.306	-92.600	22.420	-92.623	-92.444	-92.808	20.988
162	Ala	-94.061	-93.958	-94.166	18.613	-94.089	-93.929	-94.252	17.969
163	Asn	-93.589	-93.427	-93.757	22.990	-93.623	-93.492	-93.756	22.101
164	Val	-92.909	-92.775	-93.047	19.059	-92.865	-92.705	-93.030	19.635
165	Asp	-92.807	-92.709	-92.907	19.546	-92.853	-92.701	-93.008	20.284
167	Val	-93.805	-93.656	-93.958	21.963	-93.753	-93.573	-93.937	22.237
168	Phe	-92.972	-92.764	-93.190	20.653	-92.896	-92.735	-93.062	21.343
170	Asp	-93.932	-93.768	-94.102	20.974	-93.941	-93.789	-94.098	20.713
171	Leu	-94.347	-94.170	-94.529	22.238	-94.334	-94.183	-94.488	21.185

Table C.3: $^{1}J_{NH}$ of RalAΔC-GDP from $^{1}J_{NH}$-modulated TROSY-HSQC experiments

#	Resi	ns = 64*, Δ = 32.88–58.48 ms				ns = 48*, Δ = 22.24–47.84 ms			
		$^{1}J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C (s^{-1})	$^{1}J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C (s^{-1})
172	Met	-93.916	-93.764	-94.073	23.342	-93.775	-93.638	-93.914	22.261
176	Arg	-93.116	-92.953	-93.284	22.250	-93.055	-92.824	-93.295	22.714
177	Ala	-93.288	-93.144	-93.437	20.339	-93.355	-93.213	-93.501	21.168
179	Lys	-93.219	-92.988	-93.463	22.720	-93.168	-92.867	-93.483	21.729
180	Met	-93.052	-92.901	-93.209	18.617	-93.105	-92.894	-93.323	19.503
181	Glu	-92.982	-92.849	-93.119	15.983	-93.015	-92.886	-93.147	16.706
183	Ser	-92.807	-92.751	-92.864	13.058	-92.786	-92.702	-92.870	13.077
184	Lys	-92.435	-92.367	-92.504	4.145	-92.448	-92.378	-92.520	4.143

C.2 OmpX in DPC

Table C.4: $^{1}J_{NH}$ of OmpX in DPC from SOFAST-IPAP-HSQC and $^{1}J_{NH}$-modulated HSQC experiments
Conditions: 995 μM ^{15}N-OmpX, 20 mM sodium phosphate (pH 6.8), 100 mM NaCl, 5 mM EDTA, 3.8% DPC, 10% D_2O, 318 K.
SNR (Add) and SNR (Sub) are the signal-to-noise ratios of peaks in the sub-spectrum obtained by taking the sum or the difference, respectively, of the in-phase (IP) and antiphase (AP) spectra. C is the decay rate of the ^{15}N magnetisation as a result of transverse relaxation. $J_{2.5\%}$ and $J_{97.5\%}$ denote the 2.5% and 97.5% confidence intervals for the fitted J value, respectively.

Table C.4: $^{1}J_{NH}$ of OmpX in DPC from SOFAST-IPAP-HSQC and $^{1}J_{NH}$-modulated HSQC experiments

#	Resi	SOFAST-IPAP-HSQC ns = 64			$^{1}J_{NH}$-modulated HSQC ns = 40, Δ = 11.60–37.20 ms				$^{1}J_{NH}$-modulated HSQC ns = 96, Δ = 11.60–37.20 ms				$^{1}J_{NH}$-modulated TROSY-HSQC ns = 40, Δ = 11.60–37.20 ms			
		$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	$J_{2.5\%}$	$J_{97.5\%}$	C	$^{1}J_{NH}$ (Hz)	$J_{2.5\%}$	$J_{97.5\%}$	C	$^{1}J_{NH}$ (Hz)	$J_{2.5\%}$	$J_{97.5\%}$	C
2	Thr	-94.453	26.144	11.115												
3	Ser	-92.860	41.934	15.843	-93.012	-91.556	-94.674	50.485	-92.057	-91.408	-92.746	51.839	-93.082	-91.372	-95.089	50.978
4	Thr	-94.775	44.505	17.563												
6	Thr	-98.166	83.921	36.839									-92.791	-91.898	-93.786	49.578
7	Gly	-93.451	57.930	20.330					-93.822	-92.798	-94.937	57.134	-93.741	-92.441	-95.208	52.460
8	Gly	-94.956	78.884	29.624	-94.482	-93.793	-95.212	49.033	-94.003	-93.402	-94.640	52.135	-93.823	-92.858	-94.886	49.588
9	Tyr	-90.422	103.507	40.599	-91.084	-90.494	-91.715	51.331	-91.406	-90.915	-91.926	51.865				
10	Ala	-94.255	51.845	17.513	-93.565	-92.083	-95.244	46.617	-92.398	-91.619	-93.237	57.330	-93.894	-92.663	-95.276	56.382
11	Gln	-91.656	48.016	18.814	-93.613	-92.095	-95.368	48.652					-93.258	-92.225	-94.391	46.608
12	Ser	-92.581	39.633	15.903												
13	Asp	-92.009	37.811	14.418												
14	Ala	-91.669	86.342	31.441	-93.092	-92.558	-93.654	52.282	-92.733	-92.360	-93.122	45.394	-92.777	-92.187	-93.402	46.505
15	Gln	-94.739	106.997	59.268					-92.121	-91.885	-92.365	46.036	-92.237	-91.848	-92.642	47.901
21	Met	-93.110	56.251	21.795												
23	Gly	-90.191	31.611	17.819												

Table C.4: $^1J_{NH}$ of OmpX in DPC from SOFAST-IPAP-HSQC and $^1J_{NH}$-modulated HSQC experiments

#	Resi	SOFAST-IPAP-HSQC ns = 64			$^1J_{NH}$-modulated HSQC ns = 40, Δ = 11.60–37.20 ms				$^1J_{NH}$-modulated HSQC ns = 96, Δ = 11.60–37.20 ms				$^1J_{NH}$-modulated TROSY-HSQC ns = 40, Δ = 11.60–37.20 ms			
		$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^1J_{NH}$ (Hz)	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$ (Hz)	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$ (Hz)	$J_{2.5\%}$	$J_{97.5\%}$	C
24	Phe	-93.435	64.780	21.229	-94.473	-93.318	-95.736	58.534	-93.947	-92.970	-95.012	57.732	-94.115	-92.877	-95.499	56.209
25	Asn	-92.758	79.522	29.138	-91.987	-91.144	-92.903	57.700	-91.943	-91.364	-92.555	50.768	-92.206	-91.472	-93.000	52.730
26	Leu	-92.397	85.087	25.150	-92.855	-91.697	-94.172	52.494	-92.888	-92.244	-93.575	49.071	-92.901	-92.276	-93.566	50.759
27	Lys	-93.855	36.987	12.315												
28	Tyr	-90.067	63.503	24.176	-91.614	-90.536	-92.822	55.118	-91.368	-90.104	-92.831	52.861	-91.443	-90.462	-92.547	49.047
29	Arg	-97.075	47.959	16.197					-93.476	-92.353	-94.715	55.446	-93.869	-92.761	-95.087	53.709
30	Tyr	-93.076	35.443	11.098												
31	Glu	-92.581	50.038	16.309	-94.307	-93.269	-95.438	54.684	-94.652	-93.683	-95.704	49.636	-93.961	-92.611	-95.466	49.863
32	Glu	-93.510	79.221	28.680	-92.064	-91.354	-92.834	42.426	-92.156	-91.795	-92.532	45.678	-92.621	-91.635	-93.733	37.414
33	Asp	-93.380	51.846	24.938	-94.210	-92.922	-95.650	54.870	-93.519	-92.580	-94.544	49.201	-91.918	-90.377	-93.741	43.254
35	Ser	-92.252	144.003	58.155	-93.277	-92.907	-93.661	40.091	-93.167	-92.997	-93.339	39.969	-93.205	-92.862	-93.562	39.027
37	Leu	-95.901	104.511	33.276	-91.774	-91.288	-92.287	51.552	-92.117	-91.720	-92.530	53.291	-91.822	-91.235	-92.449	50.550
38	Gly	-94.605	64.083	27.587	-95.033	-94.187	-95.944	47.459	-94.307	-93.892	-94.738	51.454	-94.401	-93.457	-95.435	47.163
39	Val	-91.979	67.063	25.486	-93.189	-92.237	-94.230	57.166	-91.976	-91.347	-92.643	53.506	-92.100	-90.972	-93.363	52.339
40	Ile	-93.498	80.048	24.439	-93.383	-92.753	-94.052	51.015	-92.737	-92.133	-93.379	52.365	-94.102	-93.275	-94.997	53.279
41	Gly	-94.448	48.683	14.465												
42	Ser	-91.700	78.474	25.738	-91.402	-90.524	-92.372	49.241	-92.490	-91.800	-93.227	57.575	-91.452	-90.063	-93.073	48.924
43	Phe	-93.197	65.261	16.571					-92.923	-92.064	-93.848	51.631	-92.149	-91.283	-93.095	51.052
44	Thr	-93.556	56.980	18.261					-93.699	-92.846	-94.613	55.365	-92.917	-91.600	-94.414	49.361
45	Tyr	-95.981	78.354	22.009	-93.278	-92.279	-94.368	59.695	-92.103	-91.187	-93.116	54.485	-93.077	-92.235	-93.991	60.040
47	Glu	-92.084	36.172	13.906												
50	Arg	-95.608	28.173	17.097					-92.971	-91.769	-94.308	53.885				
53	Ser	-92.157	15.397	11.727												
54	Ser	-95.809	23.745	11.552												
57	Tyr	-92.213	81.816	53.085	-92.493	-92.187	-92.810	39.027	-92.354	-92.202	-92.509	37.139	-92.502	-92.150	-92.869	36.362

Table C.4: $^1J_{NH}$ of OmpX in DPC from SOFAST-IPAP-HSQC and $^1J_{NH}$-modulated HSQC experiments

#	Resi	SOFAST-IPAP-HSQC ns = 64			$^1J_{NH}$-modulated HSQC ns = 40, Δ = 11.60–37.20 ms				$^1J_{NH}$-modulated HSQC ns = 96, Δ = 11.60–37.20 ms				$^1J_{NH}$-modulated TROSY-HSQC ns = 40, Δ = 11.60–37.20 ms			
		$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^1J_{NH}$ (Hz)	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$ (Hz)	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$ (Hz)	$J_{2.5\%}$	$J_{97.5\%}$	C
60	Asn	-94.556	47.494	30.220									-93.889	-92.675	-95.250	44.723
61	Gln	-94.360	52.213	19.862												
62	Tyr	-89.880	43.288	19.995												
63	Tyr	-92.829	66.659	20.432	-93.137	-91.896	-94.521	53.860	-92.829	-92.275	-93.414	51.646	-92.714	-91.089	-94.604	50.227
64	Gly	-94.118	36.875	18.066	-92.972	-91.438	-94.737	45.947	-94.669	-93.550	-95.884	51.905	-94.939	-92.977	-97.208	42.915
65	Ile	-94.553	85.470	24.649	-93.242	-92.251	-94.324	52.057	-93.147	-92.389	-93.962	55.006	-92.870	-92.091	-93.713	53.871
66	Thr	-94.320	79.242	29.744	-94.537	-93.805	-95.318	52.807	-94.283	-93.789	-94.799	54.807	-94.570	-93.749	-95.452	51.117
67	Ala	-93.166	92.068	31.511					-92.357	-91.764	-92.986	52.805				
68	Gly	-94.827	80.181	31.580	-95.539	-95.023	-96.076	56.337	-94.374	-94.077	-94.679	54.111	-94.979	-94.198	-95.810	52.743
70	Ala	-94.074	92.695	37.702	-92.213	-91.784	-92.664	43.800	-92.252	-92.044	-92.466	41.115	-92.298	-91.787	-92.842	43.098
71	Tyr	-94.852	49.664	17.232	-92.546	-91.293	-93.951	57.818	-92.244	-91.276	-93.303	52.566				
72	Arg	-92.613	40.372	12.276												
74	Asn	-92.761	83.513	26.184	-92.390	-91.278	-93.629	55.215	-92.530	-91.915	-93.184	56.426	-91.737	-90.564	-93.062	56.764
75	Asp	-97.255	38.068	19.455	-92.981	-90.889	-95.521	44.087	-92.685	-91.235	-94.341	55.049				
76	Trp	-93.785	107.258	37.245					-92.819	-91.035	-95.035	43.231	-93.006	-92.303	-93.765	46.393
77	Ala	-93.721	92.260	25.137	-93.087	-91.772	-94.572	58.797	-92.364	-91.893	-92.858	54.091	-92.534	-91.680	-93.465	57.737
78	Ser	-92.087	62.329	25.122					-91.533	-90.661	-92.490	50.928	-91.941	-90.649	-93.436	53.972
79	Ile	-92.876	81.948	25.489	-91.790	-90.826	-92.856	57.724	-92.108	-91.644	-92.594	58.162	-92.045	-91.162	-93.007	55.163
81	Gly	-92.727	72.427	24.094	-92.979	-91.689	-94.443	55.888	-92.471	-91.627	-93.391	54.768	-92.960	-91.879	-94.161	52.404
82	Val								-93.042	-92.676	-93.423	51.672	-93.526	-93.016	-94.061	51.921
83	Val	-93.593	67.307	21.798	-92.788	-91.689	-94.001	57.341	-92.122	-90.979	-93.405	53.564	-92.505	-91.086	-94.144	49.978
84	Gly	-95.542	104.005	42.630	-94.193	-93.607	-94.814	51.659	-94.353	-94.039	-94.677	53.411	-94.313	-93.820	-94.830	50.999
85	Val	-91.611	87.123	25.896	-93.017	-92.395	-93.677	53.578	-92.747	-92.188	-93.334	57.643	-92.830	-92.328	-93.358	53.353
86	Gly	-93.815	72.596	25.986	-93.388	-92.002	-94.957	54.299	-93.319	-92.567	-94.126	54.700	-92.652	-91.201	-94.346	47.980
87	Tyr	-89.894	43.756	13.628									-94.234	-92.944	-95.670	46.553

Table C.4: $^1J_{NH}$ of OmpX in DPC from SOFAST-IPAP-HSQC and $^1J_{NH}$-modulated HSQC experiments

#	Resi	SOFAST-IPAP-HSQC ns = 64			$^1J_{NH}$-modulated HSQC ns = 40, Δ = 11.60–37.20 ms				$^1J_{NH}$-modulated HSQC ns = 96, Δ = 11.60–37.20 ms				$^1J_{NH}$-modulated TROSY-HSQC ns = 40, Δ = 11.60–37.20 ms			
		$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^1J_{NH}$ (Hz)	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$ (Hz)	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$ (Hz)	$J_{2.5\%}$	$J_{97.5\%}$	C
88	Gly	-94.353	33.655	14.798												
89	Lys	-86.670	32.813	14.906									-93.556	-91.639	-95.954	50.092
91	Gln	-89.955	74.652	45.115									-91.982	-91.584	-92.396	46.691
95	Tyr	-92.458	150.336	89.634	-92.389	-92.220	-92.562	30.518	-92.582	-92.487	-92.678	30.711	-92.367	-92.152	-92.589	30.084
98	Tyr	-95.045	35.365	21.722	-91.868	-90.453	-93.540	41.166	-92.030	-91.385	-92.719	38.903	-93.315	-91.960	-94.848	48.312
103	Ser	-92.329	46.023	26.014	-92.018	-90.685	-93.536	50.555	-91.631	-91.093	-92.202	49.103	-92.440	-91.255	-93.775	43.345
104	Asp	-93.340	78.264	39.698	-93.327	-92.826	-93.855	47.342	-93.131	-92.848	-93.423	45.459	-93.241	-92.681	-93.836	44.963
105	Tyr	-92.429	55.403	18.879	-91.955	-90.700	-93.375	52.137								
106	Gly	-95.997	46.981	24.491	-93.315	-92.308	-94.415	48.830	-93.349	-92.571	-94.186	50.319	-93.908	-92.623	-95.365	53.520
107	Phe	-92.296	136.741	44.547	-92.577	-92.187	-92.984	47.522	-92.444	-92.124	-92.776	47.653	-92.716	-92.238	-93.219	47.387
108	Ser	-91.791	81.591	34.344	-92.365	-91.669	-93.116	47.810	-92.772	-92.357	-93.205	45.991	-92.814	-92.110	-93.575	49.567
109	Tyr	-97.368	102.867	37.459	-93.903	-93.373	-94.460	54.714	-93.861	-93.644	-94.083	54.481	-93.631	-93.020	-94.281	53.346
110	Gly	-93.775	76.922	30.554	-94.002	-93.045	-95.049	53.335	-93.626	-92.762	-94.560	54.077	-94.183	-93.387	-95.037	54.617
111	Ala	-92.672	145.094	46.048	-92.022	-91.462	-92.616	51.690	-92.048	-91.770	-92.335	51.976	-92.190	-91.770	-92.631	50.174
112	Gly	-93.819	91.430	36.218	-93.957	-93.488	-94.447	51.847	-94.112	-93.606	-94.644	50.758	-94.155	-93.511	-94.842	49.118
113	Leu	-91.871	110.192	40.729												
114	Gln	-92.067	71.718	27.008	-92.463	-91.771	-93.207	49.683	-92.248	-91.721	-92.801	50.117	-93.234	-92.200	-94.381	50.223
115	Phe	-92.474	53.006	14.742	-93.750	-91.525	-96.409	57.534	-92.292	-91.021	-93.728	55.620	-93.324	-91.253	-95.804	56.928
116	Asn	-90.681	64.849	23.883												
118	Met	-87.178	102.258	36.855												
119	Glu	-89.949	32.702	16.867												
120	Asn	-91.865	90.390	35.814	-91.964	-91.344	-92.634	46.922	-91.652	-91.238	-92.086	47.675	-91.621	-90.790	-92.536	49.137
121	Val	-92.281	120.848	39.041	-92.369	-92.051	-92.700	47.721	-92.888	-92.581	-93.205	48.158	-92.793	-92.350	-93.258	47.474
122	Ala	-91.115	68.350	24.402	-93.073	-92.369	-93.825	50.138	-92.691	-92.136	-93.275	52.489	-93.150	-92.341	-94.024	54.550
123	Leu	-90.483	77.925	28.326	-90.634	-89.797	-91.552	52.422	-90.590	-89.946	-91.281	55.238	-90.963	-90.348	-91.622	55.269

Table C.4: $^1J_{NH}$ of OmpX in DPC from SOFAST-IPAP-HSQC and $^1J_{NH}$-modulated HSQC experiments

#	Resi	SOFAST-IPAP-HSQC ns = 64			$^1J_{NH}$-modulated HSQC ns = 40, Δ = 11.60–37.20 ms				$^1J_{NH}$-modulated HSQC ns = 96, Δ = 11.60–37.20 ms				$^1J_{NH}$-modulated TROSY-HSQC ns = 40, Δ = 11.60–37.20 ms			
		$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^1J_{NH}$ (Hz)	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$ (Hz)	$J_{2.5\%}$	$J_{97.5\%}$	C	$^1J_{NH}$ (Hz)	$J_{2.5\%}$	$J_{97.5\%}$	C
124	Asp	-92.755	74.402	23.627	-93.218	-92.430	-94.068	49.643	-92.415	-91.722	-93.160	56.751	-92.421	-91.550	-93.375	54.867
125	Phe	-94.291	106.016	34.832	-93.380	-92.662	-94.150	49.275	-93.688	-93.290	-94.102	53.165	-93.488	-92.788	-94.235	51.929
126	Ser	-95.413	94.488	30.103	-94.852	-94.245	-95.492	55.294	-95.148	-94.731	-95.580	56.367	-94.597	-93.552	-95.744	54.077
127	Tyr	-91.807	66.322	23.685	-91.798	-91.049	-92.605	55.392								
128	Glu	-93.821	49.667	15.876	-93.917	-92.303	-95.784	52.287	-93.300	-92.145	-94.579	52.868	-92.999	-91.989	-94.115	49.651
129	Gln	-92.800	51.795	17.289									-95.360	-93.506	-97.482	57.384
130	Ser	-89.725	49.125	21.703	-93.619	-92.798	-94.497	46.248	-92.749	-92.184	-93.344	48.084	-93.498	-91.929	-95.331	51.746
131	Arg	-92.697	94.538	36.631	-93.846	-93.366	-94.350	46.736	-93.633	-93.206	-94.079	45.023	-93.582	-93.123	-94.063	46.400
133	Arg	-94.124	59.092	21.274	-93.657	-92.250	-95.261	50.057	-94.282	-93.339	-95.308	55.913	-93.660	-92.676	-94.736	50.081
135	Val	-93.301	92.195	37.213	-93.952	-93.535	-94.386	44.416	-93.466	-93.073	-93.876	45.125	-93.534	-92.947	-94.159	44.873
136	Asp	-95.075	54.688	19.965	-92.889	-91.329	-94.684	50.918	-92.621	-91.810	-93.500	54.038	-93.492	-92.196	-94.958	47.136
137	Val	-92.417	58.270	25.068	-93.593	-92.701	-94.560	49.531	-92.630	-91.919	-93.394	49.790	-92.815	-92.115	-93.567	45.405
138	Gly	-92.359	59.091	22.375					-92.393	-91.579	-93.275	51.297	-93.365	-92.098	-94.783	55.164
139	Thr	-92.002	59.330	20.499	-92.664	-91.921	-93.467	49.786	-91.843	-91.118	-92.623	53.434	-93.139	-92.160	-94.208	45.989
140	Trp	-89.204	49.797	16.569					-91.729	-90.807	-92.737	53.643				
141	Ile	-91.936	43.569	15.337												
142	Ala	-92.791	83.013	23.355	-92.953	-91.841	-94.186	54.697	-93.054	-92.063	-94.137	55.579	-92.992	-92.092	-93.975	51.406
143	Gly	-95.135	82.099	34.743	-93.338	-92.742	-93.972	46.804	-94.034	-93.591	-94.496	49.867	-94.012	-93.290	-94.788	45.902
144	Val	-93.490	100.166	28.957	-92.700	-91.834	-93.646	49.663	-92.240	-91.833	-92.665	55.880	-92.571	-91.672	-93.557	52.413
145	Gly	-94.657	86.308	36.304	-94.720	-93.938	-95.552	54.112	-93.568	-93.241	-93.905	50.418	-94.235	-93.217	-95.343	53.119
146	Tyr	-93.319	67.736	27.492	-91.562	-90.471	-92.794	53.941	-91.904	-91.563	-92.257	48.250	-91.625	-90.356	-93.075	47.349
147	Arg	-91.711	34.368	14.508												
148	Phe	-92.337	62.356	21.865	-93.957	-92.630	-95.448	56.202	-92.638	-91.855	-93.480	53.272	-92.932	-91.823	-94.169	55.232

C.3 OmpX in urea

Table C.5: $^{1}J_{NH}$ of OmpX in 8 M and 6 M urea from SOFAST-IPAP-HSQC and $^{1}J_{NH}$-modulated HSQC experiments
Conditions: 221 µM ^{15}N-OmpX, 8 M urea or 166 µM ^{15}N-OmpX, 6 M urea, both in 20 mM sodium phosphate (pH 6.5), 0.1 mM NaN_3, 5% D_2O, 288 K.
SNR (Add) and SNR (Sub) are the signal-to-noise ratios of peaks in the sub-spectrum obtained by taking the sum or the difference, respectively, of the in-phase (IP) and antiphase (AP) spectra. C is the decay rate of the ^{15}N magnetisation as a result of transverse relaxation. $J_{2.5\%}$ and $J_{97.5\%}$ denote the 2.5% and 97.5% confidence intervals for the fitted *J* value, respectively.

Table C.5: $^{1}J_{NH}$ of OmpX in 8 M and 6 M urea from SOFAST-IPAP-HSQC and $^{1}J_{NH}$-modulated HSQC experiments

#	Resi	8 M urea							6 M urea		
		$^{1}J_{NH}$-modulated HSQC ns = 40, Δ = 22.24–47.84 ms				SOFAST-IPAP-HSQC ns = 64			SOFAST-IPAP-HSQC ns = 64		
		$^{1}J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^{1}J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)
2	Thr	-92.920	-92.681	-93.169	11.378	-92.848	104.694	95.760	-92.986	117.767	106.509
3	Ser	-93.116	-93.024	-93.209	4.733	-93.060	190.995	166.866	-93.133	167.895	149.431
4	Thr	-92.866	-92.809	-92.924	9.190	-92.970	206.101	130.566	-93.128	161.186	137.356
5	Val	-92.731	-92.659	-92.803	4.679	-92.708	195.983	160.457	-92.832	224.286	187.729
6	Thr	-93.225	-93.012	-93.445	4.862	-93.069	91.840	76.278	-93.152	191.164	158.420
6b	Thr								-93.255	28.013	24.515
7	Gly	-93.888	-93.814	-93.962	4.748	-93.895	239.781	194.992	-94.069	187.395	157.344
10	Ala	-93.185	-93.130	-93.241	5.234	-93.150	289.989	227.081	-93.260	252.862	197.387
11	Gln	-92.712	-92.654	-92.771	5.995	-92.701	293.187	227.012	-92.786	236.940	186.343
12	Ser	-92.904	-92.819	-92.990	6.450	-92.923	256.481	193.816	-93.016	186.947	146.937
12b	Ser								-92.967	21.490	17.252
13	Asp	-93.033	-92.963	-93.104	6.155	-92.970	270.262	203.609	-93.092	220.417	170.848
14	Ala	-93.040	-92.960	-93.122	6.080	-93.086	260.526	198.434	-93.168	213.168	166.287
14b	Ala					-92.839	45.680	34.000	-92.640	30.665	24.073

Table C.5: $^1J_{NH}$ of OmpX in 8 M and 6 M urea from SOFAST-IPAP-HSQC and $^1J_{NH}$-modulated HSQC experiments

#	Resi	8 M urea							6 M urea		
		$^1J_{NH}$-modulated HSQC ns = 40, Δ = 22.24–47.84 ms				SOFAST-IPAP-HSQC ns = 64			SOFAST-IPAP-HSQC ns = 64		
		$^1J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C	$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)
15	Gln	-92.779	-92.726	-92.831	5.212	-92.612	195.830	151.106	-92.745	167.196	129.626
15b	Gln					-92.613	31.723	34.676	-93.484	44.058	40.099
16	Gly								-94.140	132.509	107.008
17	Gln	-92.672	-92.597	-92.747	6.727	-92.568	216.967	159.588	-92.695	165.951	125.694
17b	Gln					-91.256	16.588	18.250	-92.463	38.184	35.660
17c	Gln					-92.337	23.871	24.040	-92.225	11.973	13.427
18	Met								-93.466	34.488	26.081
18b	Met								-92.726	32.717	27.309
19	Asn	-93.026	-92.885	-93.170	6.943	-92.987	148.148	111.506	-93.164	102.204	79.045
19b	Asn	-93.576	-92.917	-94.299	7.373	-92.598	31.143	22.906	-93.416	29.187	22.628
20	Lys	-92.672	-92.518	-92.829	7.740	-92.722	106.615	80.255	-92.935	71.775	55.928
20b	Lys					-93.981	15.237	18.349			
21	Met	-92.675	-92.608	-92.743	6.136	-92.602	135.081	105.342	-92.848	98.714	77.910
21b	Met					-92.613	28.840	22.162	-92.741	37.507	28.269
22	Gly	-93.874	-93.728	-94.024	6.434	-93.849	166.959	131.907	-94.031	118.549	92.882
22b	Gly	-94.036	-93.200	-94.951	10.447	-93.803	26.168	19.389	-94.564	11.409	9.472
22c	Gly					-93.556	14.729	13.602			
24	Phe	-92.623	-92.553	-92.694	7.063	-92.548	202.174	146.535	-92.739	151.716	112.405
24b	Phe					-92.320	26.780	19.050	-92.297	31.302	27.364
25	Asn	-92.953	-92.839	-93.070	8.053	-93.032	153.484	108.270	-93.146	106.335	77.761
25b	Asn					-92.965	25.495	18.861	-93.003	30.545	23.761
26	Leu	-92.784	-92.627	-92.946	8.943	-92.813	146.475	100.622	-92.941	119.374	85.227
26b	Leu	-92.616	-92.042	-93.248	11.115	-92.652	46.638	33.425	-92.811	49.430	34.343

Table C.5: $^1J_{NH}$ of OmpX in 8 M and 6 M urea from SOFAST-IPAP-HSQC and $^1J_{NH}$-modulated HSQC experiments

#	Resi	8 M urea							6 M urea		
		$^1J_{NH}$-modulated HSQC ns = 40, Δ = 22.24–47.84 ms				SOFAST-IPAP-HSQC ns = 64			SOFAST-IPAP-HSQC ns = 64		
		$^1J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C	$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)
27	Lys								-92.208	79.808	58.362
28	Tyr					-92.244	32.565	21.518	-92.859	105.717	73.400
28b	Tyr								-92.853	27.588	19.060
29	Arg	-92.565	-92.407	-92.728	12.113	-92.718	102.471	67.212	-92.758	93.587	61.416
30	Tyr	-92.559	-92.245	-92.891	9.335	-92.621	97.335	65.139	-92.713	85.928	60.995
31	Glu	-92.424	-92.281	-92.571	10.413	-92.620	108.648	73.925	-92.692	101.482	70.763
31b	Glu	-92.557	-91.676	-93.592	6.016	-92.383	34.448	26.684	-92.669	38.956	27.021
32	Glu	-92.594	-92.439	-92.752	10.624	-92.507	126.916	87.606	-92.734	112.178	78.832
32b	Glu					-93.198	43.911	36.065	-93.231	46.185	36.894
33	Asp	-92.793	-92.675	-92.914	8.796	-92.859	156.113	103.632	-92.780	143.129	99.034
34	Asn					-92.727	152.115	107.679	-92.966	138.510	100.722
35	Ser	-93.352	-93.297	-93.408	8.092	-93.259	257.748	172.462	-93.441	232.833	161.674
37	Leu								-92.753	200.167	147.342
38	Gly								-94.084	192.090	145.736
39	Val	-92.423	-92.313	-92.536	6.522	-92.393	241.023	165.384	-92.440	214.529	154.199
40	Ile	-92.897	-92.839	-92.956	7.111	-92.867	250.117	172.044	-93.001	228.280	162.328
41	Gly	-94.038	-93.962	-94.115	7.244	-93.973	224.318	163.347	-94.127	191.256	142.520
42	Ser	-92.690	-92.618	-92.763	7.766	-92.685	245.905	169.357	-92.876	192.777	136.218
43	Phe	-92.695	-92.579	-92.813	8.023	-92.716	224.755	160.106	-92.921	182.452	135.121
43b	Phe					-92.321	21.109	15.447	-92.862	28.769	22.275
43c	Phe								-92.961	24.290	23.529
44	Thr	-92.759	-92.622	-92.900	8.490	-92.707	159.755	108.663	-92.872	121.966	87.404
45	Tyr	-92.735	-92.656	-92.815	9.270	-92.777	180.853	117.498	-92.893	147.235	104.105

Table C.5: $^1J_{NH}$ of OmpX in 8 M and 6 M urea from SOFAST-IPAP-HSQC and $^1J_{NH}$-modulated HSQC experiments

#	Resi	8 M urea							6 M urea		
		$^1J_{NH}$-modulated HSQC ns = 40, Δ = 22.24–47.84 ms				SOFAST-IPAP-HSQC ns = 64			SOFAST-IPAP-HSQC ns = 64		
		$^1J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C	$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)
46	Thr	-92.724	-92.601	-92.849	10.479	-92.620	143.231	93.325	-92.952	114.472	79.442
46b	Thr					-92.271	12.450	9.053	-92.397	19.835	12.180
47	Glu	-92.726	-92.618	-92.837	9.997	-92.698	199.886	127.660	-92.849	164.665	109.237
47b	Glu					-90.849	18.981	14.532	-92.571	30.238	18.703
48	Lys	-92.718	-92.554	-92.886	8.694	-92.582	115.568	76.964	-92.794	89.702	62.246
48b	Lys					-91.040	11.219	10.579			
49	Ser	-93.040	-92.907	-93.175	9.955	-92.811	154.085	107.639	-93.208	103.023	74.317
49b	Ser					-92.564	16.391	12.600			
50	Arg	-92.614	-92.534	-92.696	10.170	-92.516	167.144	111.789	-92.776	113.109	78.752
51	Thr	-92.903	-92.779	-93.030	8.737	-92.991	160.561	113.735	-93.019	113.693	84.178
51b	Thr					-91.993	13.371	14.336	-93.109	15.756	13.412
52	Ala	-92.971	-92.869	-93.074	8.455	-92.988	213.941	151.042	-93.137	148.857	109.065
52b	Ala					-92.838	12.943	10.053			
53	Ser	-92.803	-92.715	-92.894	7.828	-92.839	194.022	145.060	-92.992	114.095	90.667
53b	Ser					-93.156	16.571	15.824	-92.893	21.305	18.574
54	Ser	-92.934	-92.855	-93.015	9.322	-92.942	194.791	123.357	-93.046	155.707	100.773
55	Gly	-94.136	-93.981	-94.294	6.915	-94.082	178.876	132.803	-94.051	121.282	93.677
56	Asp	-92.802	-92.718	-92.887	7.290	-92.819	228.885	157.231	-92.883	184.253	131.915
56b	Asp					-92.358	20.849	17.732			
57	Tyr	-92.726	-92.624	-92.830	7.231	-92.773	181.775	124.714	-92.813	154.609	108.671
57b	Tyr					-91.702	17.918	8.137			
58	Asn	-92.978	-92.883	-93.074	9.378	-93.049	138.803	94.488	-93.171	103.049	71.725
59	Lys	-92.941	-92.730	-93.159	10.939	-93.009	95.089	66.882	-92.951	74.049	53.085

Table C.5: $^1J_{NH}$ of OmpX in 8 M and 6 M urea from SOFAST-IPAP-HSQC and $^1J_{NH}$-modulated HSQC experiments

#	Resi	8 M urea							6 M urea		
		$^1J_{NH}$-modulated HSQC ns = 40, Δ = 22.24–47.84 ms				SOFAST-IPAP-HSQC ns = 64			SOFAST-IPAP-HSQC ns = 64		
		$^1J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C	$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)
59b	Lys								-92.200	33.290	29.129
60	Asn	-93.020	-92.828	-93.218	10.239	-93.141	148.161	96.873	-93.351	110.456	75.014
61	Gln	-92.584	-92.444	-92.727	7.757	-92.617	168.455	109.434	-92.805	133.107	87.230
62	Tyr	-92.485	-92.292	-92.684	8.673	-92.613	140.119	93.892	-92.662	121.118	80.353
63	Tyr					-90.725	21.351	17.756	-92.651	110.944	76.712
64	Gly	-93.957	-93.787	-94.132	7.568	-93.927	130.137	91.937	-94.145	108.092	79.711
65	Ile	-92.414	-92.303	-92.527	8.591	-92.306	194.637	130.395	-92.465	174.253	118.363
66	Thr	-92.854	-92.739	-92.970	9.208	-92.931	169.965	114.006	-93.107	128.002	90.302
66b	Thr					-92.510	16.764	13.796	-93.490	14.079	13.104
66c	Thr					-92.704	17.459	13.775			
67	Ala	-92.716	-92.628	-92.805	7.740	-92.739	212.685	149.101	-92.993	161.657	115.587
67b	Ala					-93.333	21.042	15.819	-92.636	22.606	15.471
68	Gly	-93.992	-93.877	-94.109	6.533	-93.873	222.434	169.117	-94.042	181.916	139.766
70	Ala	-92.701	-92.633	-92.770	8.809	-92.790	217.129	142.593	-92.899	173.263	117.458
71	Tyr	-92.570	-92.427	-92.716	8.486	-92.532	162.313	106.511	-92.626	131.496	89.484
72	Arg	-92.586	-92.385	-92.794	11.467	-92.505	116.070	71.095	-92.840	104.979	64.026
73	Ile	-92.625	-92.366	-92.900	10.265	-92.450	134.701	82.478	-92.716	125.638	78.198
74	Asn	-92.760	-92.538	-92.990	11.992	-92.786	122.349	74.865	-92.787	94.818	65.605
74b	Asn								-93.601	90.150	66.327
75	Asp	-92.742	-92.621	-92.865	11.359	-92.785	146.684	91.404	-92.923	124.413	82.366
76	Trp	-92.644	-92.394	-92.907	12.674	-92.487	109.692	67.713	-92.766	121.749	78.305
77	Ala	-93.177	-93.021	-93.338	10.724	-93.216	146.297	84.790	-93.135	133.294	78.967
78	Ser	-92.941	-92.764	-93.123	9.867	-92.875	144.242	86.538	-92.879	127.375	78.021

Table C.5: $^1J_{NH}$ of OmpX in 8 M and 6 M urea from SOFAST-IPAP-HSQC and $^1J_{NH}$-modulated HSQC experiments

#	Resi	8 M urea							6 M urea		
		$^1J_{NH}$-modulated HSQC ns = 40, Δ = 22.24–47.84 ms				SOFAST-IPAP-HSQC ns = 64			SOFAST-IPAP-HSQC ns = 64		
		$^1J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C	$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)
79	Ile	-92.315	-92.168	-92.466	10.414	-92.606	124.048	77.691	-92.619	118.602	75.462
80	Tyr	-92.498	-92.336	-92.664	11.656	-92.602	131.639	79.581	-92.583	98.459	65.226
81	Gly	-94.052	-93.738	-94.381	9.777	-94.080	121.511	81.228	-94.168	110.277	73.912
82	Val	-92.242	-92.095	-92.393	8.118	-92.225	196.659	128.390	-92.383	174.684	118.729
83	Val	-92.643	-92.577	-92.710	8.662	-92.765	210.798	143.968	-92.844	189.774	132.779
84	Gly	-94.084	-93.961	-94.209	7.711	-93.959	215.296	147.989	-94.075	180.831	126.479
85	Val	-92.443	-92.351	-92.537	7.380	-92.361	247.339	166.484	-92.497	212.671	146.048
86	Gly	-94.029	-93.946	-94.113	7.843	-94.009	216.556	150.675	-94.038	170.551	122.562
88	Gly	-94.080	-93.875	-94.290	7.221	-93.967	189.751	133.826	-94.170	147.934	108.990
89	Lys	-92.502	-92.347	-92.661	8.332	-92.576	128.772	85.038	-92.660	98.564	66.385
90	Phe	-92.786	-92.672	-92.903	8.335	-92.750	196.378	132.466	-92.926	154.626	104.538
91	Gln	-92.765	-92.660	-92.873	9.729	-92.581	185.759	123.827	-92.780	147.712	99.709
91b	Gln								-92.292	9.767	12.077
92	Thr	-92.857	-92.676	-93.045	9.538	-92.819	139.981	92.790	-93.141	101.765	73.236
92b	Thr	-94.148	-93.265	-95.128	8.415	-93.331	24.455	16.065	-93.401	18.972	12.590
93	Thr	-92.591	-92.432	-92.754	10.271	-92.675	133.884	83.646	-92.837	101.555	68.678
94	Glu	-92.541	-92.428	-92.656	10.726	-92.576	150.439	102.776	-92.754	123.389	85.424
94b	Glu	-92.912	-92.184	-93.733	9.614	-92.888	26.502	19.352	-92.302	21.748	14.287
95	Tyr	-92.990	-92.823	-93.161	11.178	-92.555	117.663	78.774	-92.823	123.592	81.018
97	Thr	-92.439	-92.319	-92.562	9.828	-92.483	133.007	81.106	-92.550	115.009	75.641
98	Tyr	-92.622	-92.407	-92.845	11.271	-92.543	132.077	83.386	-92.838	113.079	75.915
99	Lys	-92.899	-92.603	-93.210	13.320	-92.680	78.065	50.247	-92.907	63.722	42.259
99b	Lys					-93.262	14.337	8.555			

Table C.5: $^1J_{NH}$ of OmpX in 8 M and 6 M urea from SOFAST-IPAP-HSQC and $^1J_{NH}$-modulated HSQC experiments

#	Resi	8 M urea							6 M urea		
		$^1J_{NH}$-modulated HSQC ns = 40, Δ = 22.24–47.84 ms				SOFAST-IPAP-HSQC ns = 64			SOFAST-IPAP-HSQC ns = 64		
		$^1J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C	$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)
100	His					-92.872	58.968	40.493			
101	Asp	-92.625	-92.386	-92.876	11.435	-92.684	125.152	82.203			
102	Thr	-92.789	-92.673	-92.909	10.195	-92.709	147.974	99.574	-92.857	116.229	79.428
102b	Thr								-93.805	12.874	11.650
103	Ser	-92.840	-92.711	-92.973	9.437	-92.866	157.045	113.628	-93.171	77.908	61.285
104	Asp	-92.819	-92.720	-92.920	7.847	-92.924	197.059	128.258	-92.986	169.219	113.964
105	Tyr	-92.779	-92.659	-92.903	8.641	-92.769	201.816	137.505	-92.781	211.707	145.308
106	Gly	-94.075	-93.965	-94.187	8.097	-93.978	177.765	122.417	-94.102	153.367	111.024
107	Phe	-92.742	-92.651	-92.835	7.884	-92.765	226.065	146.800	-92.704	202.715	135.846
108	Ser	-93.059	-93.006	-93.112	6.824	-92.793	338.087	263.685	-93.013	160.476	112.702
109	Tyr	-92.856	-92.776	-92.938	8.465	-92.794	225.819	154.428	-92.927	196.898	137.943
109b	Tyr					-93.701	19.465	18.427			
110	Gly	-94.151	-94.040	-94.263	7.252	-93.977	211.742	154.262	-94.034	174.244	129.090
111	Ala	-92.884	-92.811	-92.957	6.891	-92.854	298.065	206.272	-92.897	248.149	173.975
112	Gly	-94.003	-93.913	-94.095	5.873	-94.088	244.794	185.289	-94.201	206.469	156.081
113	Leu	-92.667	-92.620	-92.715	7.573	-92.657	274.196	190.820	-92.822	247.761	175.381
113b	Leu					-92.602	20.387	12.817			
113c	Leu					-92.486	12.761	9.238			
114	Gln								-92.589	238.181	174.987
114b	Gln								-93.884	123.657	105.634
115	Phe								-92.936	199.355	141.808
116	Asn	-92.931	-92.785	-93.080	8.952	-92.991	188.279	124.497	-93.136	171.913	115.932
118	Met	-92.621	-92.496	-92.750	9.684	-92.575	184.567	114.747	-92.547	171.531	110.999

Table C.5: $^1J_{NH}$ of OmpX in 8 M and 6 M urea from SOFAST-IPAP-HSQC and $^1J_{NH}$-modulated HSQC experiments

#	Resi	8 M urea							6 M urea		
		$^1J_{NH}$-modulated HSQC ns = 40, Δ = 22.24–47.84 ms				SOFAST-IPAP-HSQC ns = 64			SOFAST-IPAP-HSQC ns = 64		
		$^1J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C	$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)
119	Glu	-92.664	-92.528	-92.802	9.597	-92.666	212.117	137.134	-92.721	193.791	129.778
119b	Glu								-92.042	10.160	14.248
120	Asn	-92.747	-92.634	-92.862	9.225	-92.687	207.873	140.764	-92.937	181.064	124.450
121	Val	-92.502	-92.411	-92.594	7.975	-92.470	212.960	144.475	-92.677	197.430	136.434
122	Ala	-93.221	-93.110	-93.333	8.138	-93.165	238.549	154.755	-93.247	221.781	148.625
123	Leu	-92.643	-92.563	-92.723	8.015	-92.576	231.837	157.577	-92.711	218.505	153.282
123b	Leu					-91.508	15.355	20.099			
124	Asp	-92.723	-92.630	-92.817	8.561	-92.664	236.352	158.660	-92.576	236.498	176.419
125	Phe	-92.578	-92.475	-92.682	9.316	-92.634	207.085	138.516	-92.778	200.559	134.802
125b	Phe								-91.779	11.729	17.911
126	Ser	-93.082	-92.958	-93.208	10.169	-93.026	199.897	121.787	-93.113	188.195	116.877
127	Tyr	-92.687	-92.581	-92.794	9.971	-92.822	261.775	162.902	-92.904	178.000	110.716
128	Glu	-92.803	-92.712	-92.896	9.477	-92.814	216.161	133.897	-92.949	210.744	133.105
129	Gln	-92.691	-92.602	-92.781	9.931	-92.838	204.150	130.846	-93.486	184.083	132.616
130	Ser	-92.992	-92.914	-93.071	5.189	-92.943	172.922	148.211	-92.995	173.911	113.731
131	Arg	-92.622	-92.519	-92.727	10.285	-92.721	195.753	125.473	-92.876	162.033	107.358
132	Ile	-92.752	-92.651	-92.855	9.455	-92.650	218.278	141.335	-92.814	200.000	131.812
132b	Ile								-92.955	14.100	8.699
132c	Ile								-91.176	10.396	12.513
133	Arg	-92.816	-92.689	-92.945	9.508	-92.851	202.650	132.383	-92.957	181.600	120.557
133b	Arg					-91.966	7.422	14.378			
134	Ser	-92.811	-92.733	-92.889	8.675	-92.839	223.380	151.891	-92.953	171.025	117.140
134b	Ser	-93.145	-92.781	-93.533	5.799	-93.074	68.609	61.163			

Table C.5: $^1J_{NH}$ of OmpX in 8 M and 6 M urea from SOFAST-IPAP-HSQC and $^1J_{NH}$-modulated HSQC experiments

#	Resi	8 M urea							6 M urea		
		$^1J_{NH}$-modulated HSQC ns = 40, Δ = 22.24–47.84 ms				SOFAST-IPAP-HSQC ns = 64			SOFAST-IPAP-HSQC ns = 64		
		$^1J_{NH}$ (Hz)	$J_{2.5\%}$ (Hz)	$J_{97.5\%}$ (Hz)	C	$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)	$^1J_{NH}$ (Hz)	SNR (Add)	SNR (Sub)
135	Val	-92.375	-92.248	-92.505	8.892	-92.289	225.898	151.070	-92.480	207.449	144.085
136	Asp	-92.962	-92.864	-93.061	8.729	-92.979	266.022	179.193	-93.047	243.071	169.583
138	Gly	-94.112	-94.025	-94.201	7.468	-94.096	231.963	160.876	-94.141	204.585	144.843
139	Thr	-92.777	-92.697	-92.858	7.116	-92.635	227.327	156.934	-92.767	196.214	136.831
140	Trp	-92.968	-92.850	-93.088	6.824	-92.795	250.070	177.648	-92.906	219.224	152.445
140b	Trp								-94.001	13.313	12.743
141	Ile	-92.791	-92.656	-92.929	8.945	-92.782	187.218	131.471	-92.879	167.240	119.087
142	Ala	-93.376	-93.300	-93.454	7.357	-93.387	282.246	196.252	-93.518	253.098	178.252
143	Gly	-93.752	-93.638	-93.869	5.421	-93.826	229.170	177.311	-93.927	187.338	143.750
144	Val	-92.479	-92.390	-92.569	5.694	-92.408	269.192	195.896	-92.461	235.834	172.912
145	Gly	-93.924	-93.840	-94.010	5.595	-93.973	252.670	193.749	-94.031	207.622	160.202
146	Tyr	-92.724	-92.669	-92.780	5.748	-92.683	291.632	218.242	-92.784	254.561	192.169
147	Arg	-92.728	-92.662	-92.794	5.765	-92.683	261.879	203.471	-92.750	223.892	175.991
147b	Arg					-93.351	28.288	18.901			
148	Phe	-92.236	-92.165	-92.307	4.151	-92.223	290.485	230.290	-92.276	261.603	207.205

Bibliography

1. Bednarska NG, Schymkowitz J, Rousseau F, Van Eldere J (2013) Protein aggregation in bacteria: the thin boundary between functionality and toxicity. *Microbiology* **159**: 1795–1806.
2. Rao RV, Bredesen DE (2004) Misfolded proteins, endoplasmic reticulum stress and neurodegeneration. *Curr Opin Cell Biol* **16**: 653–662.
3. Bowie JU (2005) Solving the membrane protein folding problem. *Nature* **438**: 581–589.
4. Anfinsen CB, Haber E, Sela M, White FH (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A* **47**: 1309–1314.
5. Dill KA (1990) Dominant forces in protein folding. *Biochemistry* **29**: 7133–7155.
6. Fersht A (1999) *Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding.* W. H. Freeman and Company.
7. Levinthal C (1968) Are there pathways for protein folding? *J Chim Phys* **65**: 44–45.
8. England JL, Haran G (2011) Role of solvation effects in protein denaturation: from thermodynamics to single molecules and back. *Annu Rev Phys Chem* **62**: 257–277.
9. Sherman E, Haran G (2006) Coil-globule transition in the denatured state of a small protein. *Proc Natl Acad Sci U S A* **103**: 11539–11543.
10. Otzen D (2011) Protein-surfactant interactions: A tale of many states. *Biochim Biophys Acta - Proteins Proteomics* **1814**: 562–591.
11. Reynolds JA, Tanford C (1970) The gross conformation of protein-sodium dodecyl sulfate complexes. *J Biol Chem* **245**: 5161–5165.
12. Lundahl P, Greijer E, Sandberg M, Cardell S, Eriksson KO (1986) A model for ionic and hydrophobic interactions and hydrogen-bonding in sodium dodecyl sulfate-protein complexes. *Biochim Biophys Acta (BBA)/Protein Struct Mol* **873**: 20–26.
13. Shirahama K, Tsujii K, Takagi T (1974) Free-boundary electrophoresis of sodium dodecyl sulfate-protein polypeptide complexes with special reference to SDS-polyacrylamide gel electrophoresis. *J Biochem* **75**: 309–319.
14. Guo XH, Zhao NM, Chen SH, Teixeira J (1990) Small-angle neutron scattering study of the structure of protein/detergent complexes. *Biopolymers* **29**: 335–346.
15. Andersen KK, Oliveira CL, Larsen KL, Poulsen FM, Callisen TH, Westh P, Pedersen JS, Otzen D (2009) The role of decorated SDS micelles in sub-CMC protein denaturation and association. *J Mol Biol* **391**: 207–226.
16. Pace NC, Scholtz JM, Grimsley GR (2014) Forces stabilizing proteins. *FEBS Lett* **588**: 2177–2184.
17. Tanford C (1970) Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* **24**: 1–95.

18. Stanley AM, Fleming KG (2008) The process of folding proteins into membranes: Challenges and progress. *Arch Biochem Biophys* **469**: 46–66.

19. Bartlett AI, Radford SE (2009) An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nat Struct Mol Biol* **16**: 582–588.

20. Baldwin RL (1996) On-pathway versus off-pathway folding intermediates. *Fold Des* **1**: R1–R8.

21. Kiefhaber T (1995) Kinetic traps in lysozyme folding. *Proc Natl Acad Sci U S A* **92**: 9029–9033.

22. Touchette NA, Perry KM, Matthews CR (1986) Folding of dihydrofolate reductase from *Escherichia coli*. *Biochemistry* **25**: 5445–5452.

23. Wedemeyer WJ, Welker E, Scheraga HA (2002) Proline cis-trans isomerization and protein folding. *Biochemistry* **41**: 14637–14644.

24. Wlodawer A, Svensson LA, Sjolin L, Gilliland GL (1988) Structure of phosphate-free ribonuclease A refined at 1.26 Å. *Biochemistry* **27**: 2705–2717.

25. Dodge RW, Scheraga HA (1996) Folding and unfolding kinetics of the proline-to-alanine mutants of bovine pancreatic ribonuclease A. *Biochemistry* **35**: 1548–1559.

26. Juminaga D, Wedemeyer WJ, Garduño-Júarez R, McDonald MA, Scheraga HA (1997) Tyrosyl interactions in the folding and unfolding of bovine pancreatic ribonuclease A: A study of tyrosine-to-phenylalanine mutants. *Biochemistry* **36**: 10131–10145.

27. Juminaga D, Wedemeyer WJ, Scheraga HA (1998) Proline isomerization in bovine pancreatic ribonuclease A. 1. unfolding conditions. *Biochemistry* **37**: 11614–11620.

28. Oka M, Montelione GT, Scheraga HA (1984) Chain-folding initiation structures in ribonuclease A: conformational free energy calculations on Ac-Asn-Pro-Tyr-NHMe, Ac-Tyr-Pro-Asn-NHMe, and related peptides. *J Am Chem Soc* **106**: 7959–7969.

29. Ahluwalia U, Katyal N, Deep S (2012) Models of protein folding. *J Proteins Proteomics* **3**: 85–93.

30. Ptitsyn OB (1973) [Stages in the mechanism of self-organization of protein molecules]. *Dokl Akad Nauk SSSR* **210**: 1213–1215.

31. Kim PS, Baldwin RL (1982) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu Rev Biochem* **51**: 459–489.

32. Dolgikh DA, Gilmanshin RI, Brazhnikov E V., Bychkova VE, Semisotnov G V., Venyaminov SY, Ptitsyn OB (1981) α-lactalbumin: compact state with fluctuating tertiary structure? *FEBS Lett* **136**: 311–315.

33. Ptitsyn OB (1995) How the molten globule became. *Trends Biochem Sci* **20**: 376–379.

34. Ptitsyn OB (1996) How molten is the molten globule? *Nat Struct Biol* **3**: 488–490.

35. Itzhaki LS, Otzen DE, Fersht AR (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol* **254**: 260–288.

36. Fersht AR (1995) Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc Natl Acad Sci U S A* **92**: 10869–10873.

37. Karplus M, Weaver DL (1994) Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci* **3**: 650–668.

38. Mauguen Y, Hartley RW, Dodson EJ, Dodson GG, Bricogne G, Chothia C, Jack A (1982) Molecular structure of a new family of ribonucleases. *Nature* **297**: 162–164.

39. Freund SM V, Wong KB, Fersht AR (1996) Initiation sites of protein folding by NMR analysis. *Proc Natl Acad Sci U S A* **93**: 10600–10603.

40. Bond CJ, Wong K-BB, Clarke J, Fersht AR, Daggett V (1997) Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway. *Proc Natl Acad Sci USA* **94**: 13409–13413.

41. Matouschek A, Kellis JT, Serrano L, Fersht AR (1989) Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340**: 122–126.

42. Matthews JM, Fersht AR (1995) Exploring the energy surface of protein folding by structure-reactivity relationships and engineered proteins: observation of Hammond behavior for the gross structure of the transition state and anti-Hammond behavior for structural elements for unfolding/. *Biochemistry* **34**: 6805–6814.

43. Kippen AD, Sancho J, Fersht AR (1994) Folding of barnase in parts. *Biochemistry* **33**: 3778–3786.

44. Hughson FM, Wright PE, Baldwin RL (1990) Structural characterization of a partly folded apomyoglobin intermediate. *Science (80-)* **249**: 1544–1548.

45. Loh SN, Kay MS, Baldwin RL (1995) Structure and stability of a second molten globule intermediate in the apomyoglobin folding pathway. *Proc Natl Acad Sci U S A* **92**: 5446–5450.

46. Jeng MF, Englander SW, Elöve GA, Roder H, Wand AJ (1990) Structural description of acid-denatured cytochrome c by hydrogen exchange and 2D NMR. *Biochemistry* **29**: 10433–10437.

47. Kay MS, Baldwin RL (1996) Packing interactions in the apomyoglobin folding intermediate. *Nat Struct Biol* **3**: 439–445.

48. Colón W, Elöve GA, Wakem LP, Sherman F, Roder H (1996) Side chain packing of the N- and C-terminal helices plays a critical role in the kinetics of cytochrome c folding. *Biochemistry* **35**: 5538–5549.

49. Otzen DE, Itzhaki LS, elMasry NF, Jackson SE, Fersht AR (1994) Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. *Proc Natl Acad Sci U S A* **91**: 10422–10425.

50. Matouschek A, Otzen DE, Itzhaki LS, Jackson SE, Fersht AR (1995) Movement of the position of the transition state in protein folding. *Biochemistry* **34**: 13656–13662.

51. Englander SW, Mayne L (2017) The case for defined protein folding pathways. *Proc Natl Acad Sci U S A* **114**: 8253–8258.

52. Frauenfelder H, Sligar S, Wolynes P (1991) The energy landscapes and motions of proteins. *Science (80-)* **254**: 1598–1603.

53. Dill KA, Chan HS (1997) From levinthal to pathways to funnels. *Nat Struct Biol* **4**: 10–19.

54. Wolynes PG, Onuchic JN, Thirumalai D (1995) Navigating the folding routes. *Science (80-)* **267**: 1619–1620.

55. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins Struct Funct Bioinforma* **21**: 167–195.

56. Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND (1995) Toward an outline of the topography of a realistic protein-folding funnel. *Proc Natl Acad Sci U S A* **92**: 3626–3630.

57. Karplus M (2011) Behind the folding funnel diagram. *Nat Chem Biol* **7**: 401–404.

58. Onuchic JN, Socci ND, Luthey-Schulten Z, Wolynes PG (1996) Protein folding funnels: The nature of the transition state ensemble. *Fold Des* **1**: 441–450.

59. Pirchi M, Ziv G, Riven I, Cohen SS, Zohar N, Barak Y, Haran G (2011) Single-molecule fluorescence spectroscopy maps the folding landscape of a large protein. *Nat Commun* **2**: 493.

60. Woodside MT, Block SM (2014) Reconstructing folding energy landscapes by single-molecule force spectroscopy. *Annu Rev Biophys* **43**: 19–39.

61. Banerjee PR, Deniz AA (2014) Shedding light on protein folding landscapes by single-molecule fluorescence. *Chem Soc Rev* **43**: 1172–1188.

62. Fedorov AN, Baldwin TO (1997) Cotranslational protein folding. *J Biol Chem* **272**: 32715–32718.

63. Komar AA (2009) A pause for thought along the co-translational folding pathway. *Trends Biochem Sci* **34**: 16–24.

64. Rodnina M V. (2016) The ribosome in action: Tuning of translational efficiency and protein folding. *Protein Sci* **25**: 1390–1406.

65. Qian W, Yang JR, Pearson NM, Maclean C, Zhang J (2012) Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet* **8**: e1002603.

66. Gardin J, Yeasmin R, Yurovsky A, Cai Y, Skiena S, Futcher B (2014) Measurement of average decoding rates of the 61 sense codons *in vivo*. *Elife* **3**: eLife.03735.

67. Dana A, Tuller T (2014) The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res* **42**: 9171–9181.

68. Wohlgemuth I, Brenner S, Beringer M, Rodnina M V. (2008) Modulation of the rate of peptidyl transfer on the ribosome by the nature of substrates. *J Biol Chem* **283**: 32229–32235.

69. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo*. *Nature* **505**: 701–705.

70. Pop C, Rouskin S, Ingolia NT, Han L, Phizicky EM, Weissman JS, Koller D (2014) Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol Syst Biol* **10**: 770.

71. Gorochowski TE, Ignatova Z, Bovenberg RAL, Roubos JA (2015) Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate. *Nucleic Acids Res* **43**: 3022–3032.

72. Lu J, Kobertz WR, Deutsch C (2007) Mapping the electrostatic potential within the ribosomal exit tunnel. *J Mol Biol* **371**: 1378–1391.

73. Lu J, Deutsch C (2008) Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J Mol Biol* **384**: 73–86.

74. Charneski CA, Hurst LD (2013) Positively charged residues are the major determinants of ribosomal velocity. *PLoS Biol* **11**: e1001508.

75. Hingorani KS, Gierasch LM (2014) Comparing protein folding *in vitro* and *in vivo*: Foldability meets the fitness challenge. *Curr Opin Struct Biol* **24**: 81–90.

76. Clark PL, King J (2001) A newly synthesized, ribosome-bound polypeptide chain adopts conformations dissimilar from early *in vitro* refolding intermediates. *J Biol Chem* **276**: 25411–25420.

77. Evans MS, Sander IM, Clark PL, M.S. E, I.M. S, P.L. C (2008) Co-translational folding promotes β-helix formation and avoids aggregation *in vivo*. *J Mol Biol* **383**: 683–692.

78. Frydman J, Erdjument-Bromage H, Tempst P, Ulrich Hartl F (1999) Co-translational domain folding as the structural basis for the rapid *de novo* folding of firefly luciferase. *Nat Struct Biol* **6**: 697–705.

79. Kaiser CM, Goldman DH, Chodera JD, Tinoco I, Bustamante C (2011) The ribosome modulates nascent protein folding. *Science (80-)* **334**: 1723–1727.

80. Hsu S-TD, Fucini P, Cabrita LD, Launay H, Dobson CM, Christodoulou J (2007) Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. *Proc Natl Acad Sci* **104**: 16516–16521.

81. Cabrita LD, Cassaignau AME, Launay HMM, Waudby CA, Wlodarski T, Camilloni C, Karyadi ME, Robertson AL, Wang X, Wentink AS, et al. (2016) A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding. *Nat Struct Mol Biol* **23**: 278–285.

82. Kramer G, Boehringer D, Ban N, Bukau B (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat Struct Mol Biol* **16**: 589–597.

83. Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Ulrich Hartl F (2013) Molecular chaperone functions in protein folding and proteostasis. *Annu Rev Biochem* **82**: 323–355.

84. White SH, Ladokhin AS, Jayasinghe S, Hristova K (2001) How membranes shape protein structure. *J Biol Chem* **276**: 32395–32398.

85. Ulloa-Aguirre A, Michael Conn P (2011) Pharmacoperones: a new therapeutic

approach for diseases caused by misfolded G protein-coupled receptors. *Recent Pat Endocr Metab Immune Drug Discov* **5**: 13–24.

86. Popot JL, Engelman DM (1990) Membrane protein folding and oligomerization: the two-stage model. *Biochemistry* **29**: 4031–4037.

87. Huang KS, Bayley H, Liao MJ, London E, Khorana HG (1981) Refolding of an integral membrane protein. Denaturation, renaturation, and reconstitution of intact bacteriorhodopsin and two proteolytic fragments. *J Biol Chem* **256**: 3802–3809.

88. Surrey T, Jahnig F (1992) Refolding and oriented insertion of a membrane protein into a lipid bilayer. *Proc Natl Acad Sci U S A* **89**: 7457–7461.

89. Tamm LK, Hong H, Liang B (2004) Folding and assembly of β-barrel membrane proteins. *Biochim Biophys Acta - Biomembr* **1666**: 250–263.

90. McMorran LM, Brockwell DJ, Radford SE (2015) Mechanistic studies of the biogenesis and folding of outer membrane proteins *in vitro* and *in vivo*: What have we learned to date? *Arch Biochem Biophys* **564**: 265–280.

91. Tafer H, Hiller S, Hilty C, Fernández C, Wüthrich K (2004) Nonrandom structure in the urea-unfolded *Escherichia coli* outer membrane protein X (OmpX). *Biochemistry* **43**: 860–869.

92. Kleinschmidt JH, Tamm LK (1999) Time-resolved distance determination by tryptophan fluorescence quenching: Probing intermediates in membrane protein folding. *Biochemistry* **38**: 4996–5005.

93. Kleinschmidt JH, Den Blaauwen T, Driessen AJM, Tamm LK (1999) Outer membrane protein A of *Escherichia coli* inserts and folds into lipid bilayers by a concerted mechanism. *Biochemistry* **38**: 5006–5016.

94. Kleinschmidt JH, Tamm LK (2002) Secondary and tertiary structure formation of the β-barrel membrane protein OmpA is synchronized and depends on membrane thickness. *J Mol Biol* **324**: 319–330.

95. Hong H, Tamm LK (2004) Elastic coupling of integral membrane protein stability to lipid bilayer forces. *Proc Natl Acad Sci U S A* **101**: 4065–4070.

96. Bos MP, Tommassen J (2004) Biogenesis of the Gram-negative bacterial outer membrane. *Curr Opin Microbiol* **7**: 610–616.

97. Selkrig J, Leyton DL, Webb CT, Lithgow T (2014) Assembly of β-barrel proteins into bacterial outer membranes. *Biochim Biophys Acta - Mol Cell Res* **1843**: 1542–1550.

98. Höhr AIC, Straub SP, Warscheid B, Becker T, Wiedemann N (2015) Assembly of β-barrel proteins in the mitochondrial outer membrane. *Biochim Biophys Acta - Mol Cell Res* **1853**: 74–88.

99. Schäfer U, Beck K, Müller M (1999) Skp, a molecular chaperone of Gram-negative bacteria, is required for the formation of soluble periplasmic intermediates of outer membrane proteins. *J Biol Chem* **274**: 24567–24575.

100. Lazar SW, Kolter R (1996) SurA assists the folding of *Escherichia coli* outer membrane proteins. *J Bacteriol* **178**: 1770–1773.

101. Ryan MT, Wagner R, Pfanner N (2000) The transport machinery for the import of

preproteins across the outer mitochondrial membrane. *Int J Biochem Cell Biol* **32**: 13–21.

102. Webb CT, Gorman MA, Lazarou M, Ryan MT, Gulbis JM (2006) Crystal structure of the mitochondrial chaperone TIM9•10 reveals a six-bladed α-propeller. *Mol Cell* **21**: 123–133.

103. Beverly KN, Sawaya MR, Schmid E, Koehler CM (2008) The Tim8-Tim13 complex has multiple substrate binding sites and binds cooperatively to Tim23. *J Mol Biol* **382**: 1144–1156.

104. Weinhäupl K, Lindau C, Hessel A, Wang Y, Schütze C, Jores T, Melchionda L, Schönfisch B, Kalbacher H, Bersch B, et al. (2018) Structural basis of membrane protein chaperoning through the mitochondrial intermembrane space. *Cell* **175**: 1365–1379.

105. Nagy JK, Lonzer WL, Sanders CR (2001) Kinetic study of folding and misfolding of diacylglycerol kinase in model membranes. *Biochemistry* **40**: 8971–8980.

106. Yang C, Horn R, Paulsen H (2003) The light-harvesting chlorophyll a/b complex can be reconstituted *in vitro* from its completely unfolded apoprotein. *Biochemistry* **42**: 4527–4533.

107. London E, Khorana HG (1982) Denaturation and renaturation of bacteriorhodopsin in detergents and lipid-detergent mixtures. *J Biol Chem* **257**: 7003–7011.

108. Otzen DE (2003) Folding of DsbB in mixed micelles: a kinetic analysis of the stability of a bacterial membrane protein. *J Mol Biol* **330**: 641–649.

109. Di Bartolo N, Compton ELR, Warne T, Edwards PC, Tate CG, Schertler GFX, Booth PJ (2016) Complete reversible refolding of a G-protein coupled receptor on a solid support. *PLoS One* **11**: e0151582.

110. Sanders MR, Findlay HE, Booth PJ (2018) Lipid bilayer composition modulates the unfolding free energy of a knotted α-helical membrane protein. *Proc Natl Acad Sci U S A* **115**: E1799–E1808.

111. Lau FW, Bowie JU (1997) A method for assessing the stability of a membrane protein. *Biochemistry* **36**: 5884–5892.

112. Barrera FN, Renart ML, Molina ML, Poveda JA, Encinar JA, Fernández AM, Neira JL, González-Ros JM (2005) Unfolding and refolding *in vitro* of a tetrameric, α-helical membrane protein: the prokaryotic potassium channel KcsA. *Biochemistry* **44**: 14344–14352.

113. Findlay HE, Rutherford NG, Henderson PJF, Booth PJ (2010) Unfolding free energy of a two-domain transmembrane sugar transport protein. *Proc Natl Acad Sci U S A* **107**: 18451–18456.

114. Miller D, Charalambous K, Rotem D, Schuldiner S, Curnow P, Booth PJ (2009) *In vitro* unfolding and refolding of the small multidrug transporter EmrE. *J Mol Biol* **393**: 815–832.

115. Roman EA, Argüello JM, González Flecha FL (2010) Reversible unfolding of a thermophilic membrane protein in phospholipid/detergent mixed micelles. *J Mol Biol* **397**: 550–559.

116. Curnow P, Booth PJ (2007) Combined kinetic and thermodynamic analysis of α-helical membrane protein unfolding. *Proc Natl Acad Sci U S A* **104**: 18970–18975.

117. Curnow P, Booth PJ (2009) The transition state for integral membrane protein folding. *Proc Natl Acad Sci U S A* **106**: 773–778.

118. Schlebach JP, Woodall NB, Bowie JU, Park C (2014) Bacteriorhodopsin folds through a poorly organized transition state. *J Am Chem Soc* **136**: 16574–16581.

119. Otzen DE (2002) Protein unfolding in detergents: effect of micelle structure, ionic strength, pH, and temperature. *Biophys J* **83**: 2219–2230.

120. Booth PJ, Flitsch SL, Stern LJ, Greenhalgh DA, Kim PS, Khorana HG (1995) Intermediates in the folding of the membrane protein bacteriorhodopsin. *Nat Struct Biol* **2**: 139–143.

121. Lemmon MA, Treutlein HR, Adams PD, Brünger AT, Engelman DM (1994) A dimerization motif for transmembrane α-helices. *Nat Struct Biol* **1**: 157–163.

122. Lemmon MA, Engelman DM (1994) Specificity and promiscuity in membrane helix interactions. *FEBS Lett* **346**: 17–20.

123. Senes A, Ubarretxena-Belandia I, Engelman DM (2001) The Cα—H···O hydrogen bond: A determinant of stability and specificity in transmembrane helix interactions. *Proc Natl Acad Sci U S A* **98**: 9056–9061.

124. Liao MJ, London E, Khorana HG (1983) Regeneration of the native bacteriorhodopsin structure from two chymotryptic fragments. *J Biol Chem* **258**: 9949–9955.

125. Sigrist H, Wenger RH, Kislig E, Wüthrich M (1988) Refolding of bacteriorhodopsin: Protease V8 fragmentation and chromophore reconstitution from proteolytic V8 fragments. *Eur J Biochem* **177**: 125–133.

126. Kahn TW, Engelman DM (1992) Bacteriorhodopsin can be refolded from two independently stable transmembrane helices and the complementary five-helix fragment. *Biochemistry* **31**: 6144–6151.

127. Ozawa S, Hayashi R, Masuda A, Iio T, Takahashi S (1997) Reconstitution of bacteriorhodopsin from a mixture of a proteinase V8 fragment and two synthetic peptides. *Biochim Biophys Acta - Biomembr* **1323**: 145–153.

128. Marti T (1998) Refolding of bacteriorhodopsin from expressed polypeptide fragments. *J Biol Chem* **273**: 9312–9322.

129. Hunt JF, Earnest TN, Bousche O, Kalghatgi K, Reilly K, Horvath C, Rothschild KJ, Engelman DM (1997) A biophysical study of integral membrane protein folding. *Biochemistry* **36**: 15156–15176.

130. Curnow P, Di Bartolo ND, Moreton KM, Ajoje OO, Saggese NP, Booth PJ (2011) Stable folding core in the folding transition state of an alpha-helical integral membrane protein. *Proc Natl Acad Sci U S A* **108**: 14133–14138.

131. Popot JL, Gerchman SE, Engelman DM (1987) Refolding of bacteriorhodopsin in lipid bilayers. A thermodynamically controlled two-stage process. *J Mol Biol* **198**: 655–676.

132. Riley ML, Wallace BA, Flitsch SL, Booth PJ (1997) Slow α helix formation during folding of a membrane protein. *Biochemistry* **36**: 192–196.

133. Booth PJ, Farooq A (1997) Intermediates in the assembly of bacteriorhodopsin investigated by time-resolved absorption spectroscopy. *Eur J Biochem* **246**: 674–680.

134. Booth PJ (2000) Unravelling the folding of bacteriorhodopsin. *Biochim Biophys Acta - Bioenerg* **1460**: 4–14.

135. Dutta A, Kim TY, Moeller M, Wu J, Alexiev U, Klein-Seetharaman J (2010) Characterization of membrane protein non-native states. 2. the SDS-unfolded states of rhodopsin. *Biochemistry* **49**: 6329–6340.

136. Ridge KD, Lee SS, Yao LL (1995) *In vivo* assembly of rhodopsin from expressed polypeptide fragments. *Proc Natl Acad Sci U S A* **92**: 3204–3208.

137. Sapra KT, Park PSH, Palczewski K, Muller DJ (2008) Mechanical properties of bovine rhodopsin and bacteriorhodopsin: Possible roles in folding and function. *Langmuir* **24**: 1330–1337.

138. Dutta A, Altenbach C, Mangahas S, Yanamala N, Gardner E, Hubbell WL, Klein-Seetharaman J (2014) Differential dynamics of extracellular and cytoplasmic domains in denatured states of rhodopsin. *Biochemistry* **53**: 7160–7169.

139. Tastan O, Yu E, Ganapathiraju M, Aref A, Rader AJ, Klein-Seetharaman J (2007) Comparison of stability predictions and simulated unfolding of rhodopsin structures. *Photochem Photobiol* **83**: 351–362.

140. Klein-Seetharaman J (2005) Dual role of interactions between membranous and soluble portions of helical membrane receptors for folding and signaling. *Trends Pharmacol Sci* **26**: 183–189.

141. Otzen DE (2011) Mapping the folding pathway of the transmembrane protein DsbB by protein engineering. *Protein Eng Des Sel* **24**: 139–149.

142. Gold VAM, Duong F, Collinson I (2007) Structure and function of the bacterial Sec translocon. *Mol Membr Biol* **24**: 387–394.

143. Du Plessis DJF, Nouwen N, Driessen AJM (2011) The Sec translocase. *Biochim Biophys Acta - Biomembr* **1808**: 851–865.

144. Kumazaki K, Chiba S, Takemoto M, Furukawa A, Nishiyama KI, Sugano Y, Mori T, Dohmae N, Hirata K, Nakada-Nakura Y, et al. (2014) Structural basis of Sec-independent membrane protein insertion by YidC. *Nature* **509**: 516–520.

145. Dalbey RE, Kuhn A, Zhu L, Kiefer D (2014) The membrane insertase YidC. *Biochim Biophys Acta - Mol Cell Res* **1843**: 1489–1496.

146. Shao S, Hegde RS (2011) Membrane protein insertion at the endoplasmic reticulum. *Annu Rev Cell Dev Biol* **27**: 25–56.

147. Shan SO, Stroud RM, Walter P (2004) Mechanism of association and reciprocal activation of two GTPases. *PLoS Biol* **2**: e320.

148. Andersson H, Von Heijne G (1993) Sec dependent and sec independent assembly of *E. coli* inner membrane proteins: the topological rules depend on chain length. *Eur Mol Biol Organ J* **12**: 683–691.

149. Cymer F, Von Heijne G, White SH (2015) Mechanisms of integral membrane protein insertion and folding. *J Mol Biol* **427**: 999–1022.

150. Dalbey RE, Kuhn A (2014) How YidC inserts and folds proteins across a membrane. *Nat Struct Mol Biol* **21**: 435–436.

151. Kol S, Nouwen N, Driessen AJM (2008) Mechanisms of YidC-mediated insertion and assembly of multimeric membrane protein complexes. *J Biol Chem* **283**: 31269–31273.

152. Jaud S, Fernández-Vidal M, Nilsson I, Meindl-Beinker NM, Hübner NC, Tobias DJ, von Heijne G, White SH (2009) Insertion of short transmembrane helices by the Sec61 translocon. *Proc Natl Acad Sci* **106**: 11588–11593.

153. Hedin LE, Öjemalm K, Bernsel A, Hennerdal A, Illergård K, Enquist K, Kauko A, Cristobal S, von Heijne G, Lerch-Bader M, et al. (2010) Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. *J Mol Biol* **396**: 221–229.

154. Shimohata N, Nagamori S, Akiyama Y, Kaback HR, Ito K (2007) SecY alterations that impair membrane protein folding and generate a membrane stress. *J Cell Biol* **176**: 307–317.

155. Nagamori S, Smirnova IN, Kaback HR (2004) Role of YidC in folding of polytopic membrane proteins. *J Cell Biol* **165**: 53–62.

156. Sadlish H, Pitonzo D, Johnson AE, Skach WR (2005) Sequential triage of transmembrane segments by Sec61α during biogenesis of a native multispanning membrane protein. *Nat Struct Mol Biol* **12**: 870–878.

157. Meijberg W, Booth PJ (2002) The activation energy for insertion of transmembrane α-helices is dependent on membrane composition. *J Mol Biol* **319**: 839–853.

158. Allen SJ, Curran AR, Templer RH, Meijberg W, Booth PJ (2004) Controlling the folding efficiency of an integral membrane protein. *J Mol Biol* **342**: 1293–1304.

159. Lorch M, Booth PJ (2004) Insertion kinetics of a denatured α helical membrane protein into phospholipid bilayer vesicles. *J Mol Biol* **44**: 1109–1121.

160. Miller DM, Findlay HE, Ces O, Templer RH, Booth PJ (2016) Light-activated control of protein channel assembly mediated by membrane mechanics. *Nanotechnology* **27**: 494004.

161. Harris NJ, Reading E, Ataka K, Grzegorzewski L, Charalambous K, Liu X, Schlesinger R, Heberle J, Booth PJ (2017) Structure formation during translocon-unassisted co-translational membrane protein folding. *Sci Rep* **7**: 8021.

162. Allen SJ, Curran AR, Templer RH, Meijberg W, Booth PJ (2004) Folding kinetics of an α helical membrane protein in phospholipid bilayer vesicles. *J Mol Biol* **342**: 1279–1291.

163. Seddon AM, Lorch M, Ces O, Templer RH, Macrae F, Booth PJ (2008) Phosphatidylglycerol lipids enhance folding of an alpha helical membrane protein. *J Mol Biol* **380**: 548–556.

164. Findlay HE, Booth PJ (2017) The folding, stability and function of lactose permease differ in their dependence on bilayer lipid composition. *Sci Rep* **7**: 13056.

165. Bogdanov M, Heacock PN, Dowhan W (2002) A polytopic membrane protein displays a reversible topology dependent on membrane lipid composition. *EMBO J* **21**: 2107–

2116.

166. Wang X, Bogdanov M, Dowhan W (2002) Topology of polytopic membrane protein subdomains is dictated by membrane phospholipid composition. *EMBO J* **21**: 5673–5681.

167. Gupta K, Donlan JAC, Hopper JTS, Uzdavinys P, Landreh M, Struwe WB, Drew D, Baldwin AJ, Stansfeld PJ, Robinson C V. (2017) The role of interfacial lipids in stabilizing membrane protein oligomers. *Nature* **541**: 421–424.

168. Henrich E, Peetz O, Hein C, Laguerre A, Hoffmann B, Hoffmann J, Dötsch V, Bernhard F, Morgner N (2017) Analyzing native membrane protein assembly in nanodiscs by combined non-covalent mass spectrometry and synthetic biology. *Elife* **6**: e20954.

169. Valiyaveetil FI, Zhou Y, MacKinnon R (2002) Lipids in the structure, folding, and function of the KcsA K^+ channel. *Biochemistry* **41**: 10771–10777.

170. Gold VAM, Robson A, Bao H, Romantsov T, Duong F, Collinson I (2010) The action of cardiolipin on the bacterial translocon. *Proc Natl Acad Sci*.

171. Lee AG (2004) How lipids affect the activities of integral membrane proteins. *Biochim Biophys Acta - Biomembr* **1666**: 62–87.

172. Klein N, Hellmann N, Schneider D (2015) Anionic lipids modulate the activity of the aquaglyceroporin GlpF. *Biophys J* **109**: 722–731.

173. Harris NJ, Charalambous K, Findlay HE, Booth PJ (2018) Lipids modulate the insertion and folding of the nascent chains of alpha helical membrane proteins. *Biochem Soc Trans* **46**: 1355–1366.

174. Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, Nilsson IM, White SH, Von Heijne G (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**: 377–381.

175. Heinrich SU, Mothes W, Brunner J, Rapoport TA (2000) The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell* **102**: 233–244.

176. Scotti PA, Valent QA, Manting EH, Urbanus ML, Driessen AJM, Oudega B, Luirink J (1999) SecA is not required for signal recognition particle-mediated targeting and initial membrane insertion of a nascent inner membrane protein. *J Biol Chem* **274**: 29883–29888.

177. Valent QA, Scotti PA, High S, De Gier JWL, Von Heijne G, Lentzen G, Wintermeyer W, Oudega B, Luirink J (1998) The *Escherichia coli* SRP and SecB targeting pathways converge at the translocon. *EMBO J* **17**: 2504–2512.

178. Bahari L, Parlitz R, Eitan A, Stjepanovic G, Bochkareva ES, Sinning I, Bibi E (2007) Membrane targeting of ribosomes and their release require distinct and separable functions of FtsY. *J Biol Chem* **282**: 32168–32175.

179. De Leeuw E, Te Kaat K, Moser C, Menestrina G, Demel R, De Kruijff B, Oudega B, Luirink J, Sinning I (2000) Anionic phospholipids are involved in membrane association of FtsY and stimulate its GTPase activity. *EMBO J* **19**: 531–541.

180. Lam VQ, Akopian D, Rome M, Henningsen D, Shan SO (2010) Lipid activation of the signal recognition particle receptor provides spatial coordination of protein targeting. *J Cell Biol* **190**: 623–635.

181. van Meer G, de Kroon AIPM (2011) Lipid map of the mammalian cell. *J Cell Sci* **124**: 5–8.

182. Arunan E, Desiraju GR, Klein RA, Sadlej J, Scheiner S, Alkorta I, Clary DC, Crabtree RH, Dannenberg JJ, Hobza P, et al. (2011) Definition of the hydrogen bond (IUPAC Recommendations 2011). *Pure Appl Chem* **83**: 1637–1641.

183. Desiraju GR (2011) A bond by any other name. *Angew Chemie - Int Ed* **50**: 52–59.

184. White SH, Wimley WC (1999) Membrane protein folding and stability: physical principles. *Annu Rev Biophys Biomol Struct* **28**: 319–365.

185. Dingley AJ, Cordier F, Grzesiek S (2001) An introduction to hydrogen bond scalar couplings. *Concepts Magn Reson* **13**: 103–127.

186. Persson F, Halle B (2015) How amide hydrogens exchange in native proteins. *Proc Natl Acad Sci U S A* **112**: 10383–10388.

187. Raschle T, Rios Flores P, Opitz C, Müller DJ, Hiller S (2016) Monitoring backbone hydrogen-bond formation in β-barrel membrane protein folding. *Angew Chemie - Int Ed* **55**: 5952–5955.

188. Fu Y, Gao J, Bieschke J, Dendle MA, Kelly JW (2006) Amide-to-*E*-olefin versus amide-to-ester backbone H-bond perturbations: Evaluating the O-O repulsion for extracting H-bond energies. *J Am Chem Soc* **128**: 15948–15949.

189. Gao J, Bosco DA, Powers ET, Kelly JW (2009) Localized thermodynamic coupling between hydrogen bonding and microenvironment polarity substantially stabilizes proteins. *Nat Struct Mol Biol* **16**: 684–690.

190. Edison AS, Weinhold F, Markley JL (1995) Theoretical studies of protium/deuterium fractionation factors and cooperative hydrogen bonding in peptides. *J Am Chem Soc* **117**: 9619–9624.

191. Loh SN, Markley JL (1994) Hydrogen bonding in proteins as studied by amide hydrogen D/H fractionation factors: application to staphylococcal nuclease. *Biochemistry* **33**: 1029–1036.

192. Bowers PM, Klevit RE (1996) Hydrogen bonding and equilibrium isotope enrichment in histidine-containing proteins. *Nat Struct Biol* **3**: 522–531.

193. LiWang AC, Bax A (1996) Equilibrium protium/deuterium fractionation of backbone amides in U–^{13}C/^{15}N labeled human ubiquitin by triple resonance NMR. *J Am Chem Soc* **118**: 12864–12865.

194. Khare D, Alexander P, Orban J (1999) Hydrogen bonding and equilibrium protium-deuterium fractionation factors in the immunoglobulin G binding domain of protein G. *Biochemistry* **38**: 3918–3925.

195. Krantz BA, Moran LB, Kentsis A, Sosnick TR (2000) D/H amide kinetic isotope effects reveal when hydrogen bonds form during protein folding. *Nat Struct Biol* **7**: 62–71.

196. Krantz BA, Srivastava AK, Nauli S, Baker D, Sauer RT, Sosnick TR (2002) Understanding protein hydrogen bond formation with kinetic H/D amide isotope effects. *Nat Struct Biol* **9**: 458–463.

197. LiWang AC, Bax A (1997) Solution NMR characterization of hydrogen bonds in a protein by indirect measurement of deuterium quadrupole couplings. *J Magn Reson* **127**: 54–64.

198. Marion D (2013) An introduction to biological NMR spectroscopy. *Mol Cell Proteomics* **12**: 3006–3025.

199. Keeler J (2005) *Understanding NMR Spectroscopy*. John Wiley & Sons Ltd.

200. Bystrov VF (1976) Spin-spin coupling and the conformational states of peptide systems. *Prog Nucl Magn Reson Spectrosc* **10**: 41–81.

201. Xiang SQ, Narayanan RL, Becker S, Zweckstetter M (2013) N-H spin-spin couplings: probing hydrogen bonds in proteins. *Angew Chemie - Int Ed* **52**: 3525–3528.

202. Vriend G (1990) WHAT IF: A molecular modeling and drug design program. *J Mol Graph* **8**: 52–56.

203. Juranić N, Ilich PK, Macura S (1995) Hydrogen bonding networks in proteins as revealed by the amide $^{1}J_{NC'}$ coupling constant. *J Am Chem Soc* **117**: 405–410.

204. Cornilescu G, Bax A, Case DA (2000) Large variations in one-bond $^{13}C^{\alpha}$–$^{13}C^{\beta}$ *J* couplings in polypeptides correlate with backbone conformation. *J Am Chem Soc* **122**: 2168–2171.

205. Vuister GW, Delaglio F, Bax A (1992) An empirical correlation between $^{1}J_{C\alpha H\alpha}$ and protein backbone conformation. *J Am Chem Soc* **114**: 9674–9675.

206. Karplus M (1959) Contact electro-spin coupling of nuclear magnetic moments. *J Chem Phys* **30**: 10–15.

207. Cordier F, Grzesiek S (1999) Direct observation of hydrogen bonds in proteins by interresidue $^{3h}J_{NC'}$ scalar couplings. *J Am Chem Soc* **121**: 1601–1602.

208. Xiang S, Zweckstetter M (2014) Detection of a transient intramolecular hydrogen bond using $^{1}J_{NH}$ scalar couplings. *J Magn Reson* **243**: 93–97.

209. Steinmetz MO, Jelesarov I, Matousek WM, Honnappa S, Jahnke W, Missimer JH, Frank S, Alexandrescu AT, Kammerer RA (2007) Molecular basis of coiled-coil formation. *Proc Natl Acad Sci U S A* **104**: 7062–7067.

210. Spudich JL, Yang CS, Jung KH, Spudich EN (2000) Retinylidene proteins: structures and functions from archaea to humans. *Annu Rev Cell Dev Biol* **16**: 365–392.

211. Ernst OP, Lodowski DT, Elstner M, Hegemann P, Brown LS, Kandori H (2014) Microbial and animal rhodopsins: structures, functions, and molecular mechanisms. *Chem Rev* **114**: 126–163.

212. Lanyi JK (2006) Proton transfers in the bacteriorhodopsin photocycle. *Biochim Biophys Acta - Bioenerg* **1757**: 1012–1018.

213. Hargrave PA (2001) Rhodopsin structure, function, and topography: The Friedenwald lecture. *Investig Ophthalmol Vis Sci* **42**: 3–9.

214. Luecke H, Schobert B, Richter HT, Cartailler JP, Lanyi JK (1999) Structure of bacteriorhodopsin at 1.55 Å resolution. *J Mol Biol* **291**: 899–911.

215. Okada T, Fujiyoshi Y, Silow M, Navarro J, Landau EM, Shichida Y (2002) Functional role of internal water molecules in rhodopsin revealed by X-ray crystallography. *Proc Natl Acad Sci U S A* **99**: 5982–5987.

216. Gautier A, Mott HR, Bostock MJ, Kirkpatrick JP, Nietlispach D (2010) Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. *Nat Struct Mol Biol* **17**: 768–774.

217. Inoue K, Tsukamoto T, Sudo Y (2014) Molecular and evolutionary aspects of microbial sensory rhodopsins. *Biochim Biophys Acta - Bioenerg* **1837**: 562–577.

218. Engelhard M, Scharf B, Siebert F (1996) Protonation changes during the photocycle of sensory rhodopsin II from *Natronobacterium pharaonis*. *FEBS Lett* **395**: 195–198.

219. Furutani Y, Iwamoto M, Shimono K, Wada A, Ito M, Kamo N, Kandori H (2004) FTIR spectroscopy of the O photointermediate in *pharaonis* phoborhodopsin. *Biochemistry* **83**: 3482–3489.

220. Iwamoto M, Hasegawa C, Sudo Y, Shimono K, Araiso T, Kamo N (2004) Proton release and uptake of *pharaonis* phoborhodopsin (sensory rhodopsin II) reconstituted into phospholipids. *Biochemistry* **43**: 3195–3203.

221. Schmies G, Lüttenberg B, Chizhov I, Engelhard M, Becker A, Bamberg E (2000) Sensory rhodopsin II from the haloalkaliphilic *Natronobacterium pharaonis*: Light-activated proton transfer reactions. *Biophys J* **78**: 967–976.

222. Sudo Y, Iwamoto M, Shimono K, Sumi M, Kamo N (2001) Photo-induced proton transport of *pharaonis* phoborhodopsin (sensory rhodopsin II) is ceased by association with the transducer. *Biophys J* **80**: 916–922.

223. Chizhov I, Chernavskii DS, Engelhard M, Mueller KH, Zubov B V., Hess B (1996) Spectrally silent transitions in the bacteriorhodopsin photocycle. *Biophys J* **71**: 2329–2345.

224. Chizhov I, Schmies G, Seidel R, Sydor JR, Luttenberg B, Engelhard M (1998) The photophobic receptor from *Natronobacterium pharaonis*: temperature and pH dependencies of the photocycle of sensory rhodopsin II. *Biophys J* **75**: 999–1009.

225. Gordeliy VI, Labahn J, Moukhametzianov R, Efremov R, Granzin J, Schlesinger R, Büldt G, Savopol T, Scheidig AJ, Klare JP, et al. (2002) Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex. *Nature* **419**: 484–487.

226. Falke JJ, Hazelbauer GL (2001) Transmembrane signaling in bacterial chemoreceptors. *Trends Biochem Sci* **26**: 257–265.

227. Zhuravleva A, Korzhnev DM (2017) Protein folding by NMR. *Prog Nucl Magn Reson Spectrosc* **100**: 52–77.

228. Mehler M, Eckert CE, Busche A, Kulhei J, Michaelis J, Becker-Baldus J, Wachtveitl J, Dötsch V, Glaubitz C (2015) Assembling a correctly folded and functional heptahelical membrane protein by protein *trans*-splicing. *J Biol Chem* **290**: 27712–27722.

229. Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H (2002) Long-range interactions within a nonnative protein. *Science (80-)* **295**: 1719–1722.

230. Greenfield NJ (2007) Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat Protoc* **1**: 2527–2535.

231. Chu G, Wang X, Chen T, Xu W, Wang Y, Song H, Xu Y (2015) Chiral electronic transitions of YVO_4:Eu^{3+} nanoparticles in cellulose based photonic materials with circularly polarized excitation. *J Mater Chem C* **3384**: 3384–3390.

232. Miles AJ, Wallace BA (2016) Circular dichroism spectroscopy of membrane proteins. *Chem Soc Rev* **45**: 4859–4872.

233. Sreerama N, Woody RW (2004) Computation and analysis of protein circular dichroism spectra. *Methods Enzymol* **383**: 318–351.

234. Hsu MC, Woody RW (1971) The origin of the heme Cotton effects in myoglobin and hemoglobin. *J Am Chem Soc* **93**: 3515–3525.

235. Pescitelli G, Sreerama N, Salvadori P, Nakanishi K, Berova N, Woody RW (2008) Inherent chirality dominates the visible/near-ultraviolet CD spectrum of rhodopsin. *J Am Chem Soc* **130**: 6170–6181.

236. Sreerama N, Woody RW (2004) On the analysis of membrane protein circular dichroism spectra. *Protein Sci* **13**: 100–112.

237. Abdul-Gader A, Miles AJ, Wallace BA (2011) A reference dataset for the analyses of membrane protein secondary structures and transmembrane residues using circular dichroism spectroscopy. *Bioinformatics* **27**: 1630–1636.

238. Greenfield NJ (1996) Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. *Anal Biochem* **235**: 1–10.

239. Clayden J, Greeves N, Warren S, Wothers P (2012) *Organic Chemistry*. Oxford University Press.

240. Horwitz J, Heller J (1973) Interactions of all-*trans*-, 9-, 11-, and 13-*cis*-retinal, all-*trans*-retinyl acetate, and retinoic acid with human retinol-binding protein and prealbumin. *J Biol Chem* **248**: 6317–6324.

241. Freedman KA, Becker RS (1986) Comparative investigation of the photoisomerization of the protonated and unprotonated *n*-butylamine Schiff bases of 9-*cis*-, 11-*cis*-, 13-*cis*-, and all-*trans*-retinals. *J Am Chem Soc* **108**: 1245–1251.

242. Nakanishi K, Balogh-Nair V, Amaboldi M, Tsujimoto K, Honig B (1980) An external point-charge model for bacteriorhodopsin to account for its purple color. *J Am Chem Soc* **102**: 7945–7947.

243. Takahashi T, Yan B, Spudich JL, Yan B, Mazur P, Derguini F, Nakanishi K (1990) Color regulation in the archaebacterial phototaxis receptor phoborhodopsin (sensory rhodopsin II). *Biochemistry* **29**: 8467–8474.

244. Nathans J (1990) Determinants of visual pigment absorbance: identification of the retinylidene Schiff's base counterion in bovine rhodopsin. *Biochemistry* **29**: 9746–

9752.

245. Spudich JL, Bogomolni RA (1983) Spectroscopic discrimination of the three rhodopsinlike pigments in *Halobacterium halobium* membranes. *Biophys J* **43**: 243–246.

246. Hiroaki Tomioka, Hiroyuki Sasabe (1995) Isolation of photochemically active archaebacterial photoreceptor, pharaonis phoborhodopsin from *Natronobacterium pharaonis*. *BBA - Biomembr* **1234**: 261–267.

247. Shichi H (1970) Spectrum and purity of bovine rhodopsin. *Biochemistry* **9**: 1973–1977.

248. Becher B, Cassim JY (1976) Effects of light adaptation on the purple membrane structure of *Halobacterium halobium*. *Biophys J* **16**: 1183–1200.

249. Levenberg K (1944) A method for the solution of certain non-linear problems in least squares. *Q J Appl Math* **2**: 164–168.

250. Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. *J Soc Ind Appl Math* **11**: 431–441.

251. Gavin HP (2017) The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems.

252. Billo EJ (2001) *Excel for Chemists : A Comprehensive Guide*. John Wiley & Sons, Inc.

253. Alexiev U, Farrens DL (2014) Fluorescence spectroscopy of rhodopsins: Insights and approaches. *Biochim Biophys Acta - Bioenerg* **1837**: 694–709.

254. Ghisaidoobe ABT, Chung SJ (2014) Intrinsic tryptophan fluorescence in the detection and analysis of proteins: a focus on Förster resonance energy transfer techniques. *Int J Mol Sci* **15**: 22518–22538.

255. Kandori H, Furutani Y, Nishimura S, Shichida Y, Chosrowjan H, Shibata Y, Mataga N (2001) Excited-state dynamics of rhodopsin probed by femtosecond fluorescence spectroscopy. *Chem Phys Lett* **334**: 271–276.

256. Vivian JT, Callis PR (2001) Mechanisms of tryptophan fluorescence shifts in proteins. *Biophys J* **80**: 2093–2109.

257. Fuentes L, Oyola J, Fernândez M, Quiñones E (2004) Conformational changes in azurin from *Pseudomona aeruginosa* induced through chemical and physical protocols. *Biophys J* **87**: 1873–1880.

258. Pedersen JS (2010) The nature of amyloid-like glucagon fibrils. *J Diabetes Sci Technol* **4**: 1357–1367.

259. Teale FW, Weber G (1957) Ultraviolet fluorescence of the aromatic amino acids. *Biochem J* **65**: 476–482.

260. Callis PR (1997) ^1L_a and ^1L_b transitions of tryptophan: Applications of theory and experimental observations to fluorescence of proteins. *Methods Enzymol* **278**: 113–150.

261. Hixon J, Reshetnyak YK (2009) Algorithm for the analysis of tryptophan fluorescence spectra and their correlation with protein structural parameters. *Algorithms* **2**: 1155–1176.

262. Kalisky O, Feitelson J, Ottolenghi M (1981) Photochemistry and fluorescence of

bacteriorhodopsin excited in its 280-nm absorption band. *Biochemistry* **20**: 205–209.

263. Renthal R, Brancaleon L, Peña I, Silva F, Chen LY (2011) Interaction of a two-transmembrane-helix peptide with lipid bilayers and dodecyl sulfate micelles. *Biophys Chem* **159**: 321–327.
264. Snodderly DM, Jr (1967) Reversible and irreversible bleaching of rhodopsin in detergent solutions. *Proc Natl Acad Sci U S A* **57**: 1356–1362.
265. Dancsházy Z, Tokaji Z, Dér A (1999) Bleaching of bacteriorhodopsin by continuous light. *FEBS Lett* **450**: 154–157.
266. Perrin CL (2017) Linear or nonlinear least-squares analysis of kinetic data? *J Chem Educ* **94**: 669–672.
267. Bisby RH, Thomas EW (1986) Kinetic analysis by the method of nonlinear least squares: A reaction involving consecutive steps. *J Chem Educ* **63**: 990.
268. Akaike H (1976) Canonical correlation analysis of time series and the use of an information criterion. *Math Sci Eng* **126**: 27–96.
269. Honig B, Greenberg AD, Dinur U, Ebrey TG (1976) Visual-pigment spectra: implications of the protonation of the retinal Schiff base. *Biochemistry* **15**: 4593–4599.
270. Dutta A, Tirupula KC, Alexiev U, Klein-Seetharaman J (2010) Characterization of membrane protein non-native states. 1. Extent of unfolding and aggregation of rhodopsin in the presence of chemical denaturants. *Biochemistry* **49**: 6317–6328.
271. Kiefer H, Krieger J, Olszewski JD, Von Heijne G, Prestwich GD, Breer H (1996) Expression of an olfactory receptor in *Escherichia coli*: Purification, reconstitution, and ligand binding. *Biochemistry* **35**: 16077–16084.
272. Das TK, Mazumdar S (1998) Unfolding pathway of cytochrome c oxidase induced by ionic surfactants: Circular dichroism and picosecond time-resolved fluorescence studies. *Proc Indian Acad Sci Chem Sci* **110**: 479–490.
273. Sonoda Y, Newstead S, Hu N-J, Alguel Y, Nji E, Beis K, Yashiro S, Lee C, Leung J, Cameron AD, et al. (2011) Benchmarking membrane protein detergent stability for improving throughput of high-resolution X-ray structures. *Structure* **19**: 17–25.
274. Kalipatnapu S, Chattopadhyay A (2005) Membrane protein solubilization: Recent advances and challenges in solubilization of serotonin$_{1A}$ receptors. *IUBMB Life* **57**: 505–512.
275. Arachea BT, Sun Z, Potente N, Malik R, Isailovic D, Viola RE (2012) Detergent selection for enhanced extraction of membrane proteins. *Protein Expr Purif* **86**: 12–20.
276. Chen GQ, Gouaux E (1999) Probing the folding and unfolding of wild-type and mutant forms of bacteriorhodopsin in micellar solutions: Evaluation of reversible unfolding conditions. *Biochemistry* **38**: 15380–15387.
277. Michel H (2006) 4.2. Crystallization of membrane proteins. *Int Tables Crystallogr* **F**: 94–99.
278. Yu Y (2015) Characterisation of the membrane protein complex formed by Natronomonas pharaonis sensory rhodopsin pSRII and its cognate transducer pHtrII (Ph.D. thesis).

279. Hamada H, Arakawa T, Shiraki K (2009) Effect of additives on protein aggregation. *Curr Pharm Biotechnol* **10**: 400–407.

280. Gautier A (2009) Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy (Ph.D. thesis).

281. Zagorski MG, Yang J, Shao H, Ma K, Zeng H, Hong A (1999) Methodological and chemical factors affecting amyloid β peptide amyloidogenicity. *Methods Enzymol* **309**: 189–204.

282. Chen S, Wetzel R (2001) Solubilization and disaggregation of polyglutamine peptides. *Protein Sci* **10**: 887–891.

283. Chi EY, Krishnan S, Randolph TW, Carpenter JF (2003) Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. *Pharm Res* **20**: 1325–1336.

284. Vuillard L, Rabilloud T, Goldberg ME (1998) Interactions of non-detergent sulfobetaines with early folding intermediates facilitate in vitro protein renaturation. *Eur J Biochem* **256**: 128–135.

285. Hagn F, Etzkorn M, Raschle T, Wagner G (2013) Optimized phospholipid bilayer nanodiscs facilitate high-resolution structure determination of membrane proteins. *J Am Chem Soc* **135**: 1919–1925.

286. Reddy K. RC (2005) L-Arginine increases the solubility of unfolded species of hen egg white lysozyme. *Protein Sci* **14**: 929–935.

287. Shukla D, Trout BL (2010) Interaction of arginine with proteins and the mechanism by which it inhibits aggregation. *J Phys Chem B* **114**: 13426–13438.

288. Sharma S, Sarkar S, Paul SS, Roy S, Chattopadhyay K (2013) A small molecule chemical chaperone optimizes its unfolded state contraction and denaturant like properties. *Sci Rep* **3**: 3525.

289. Breydo L, Reddy KD, Piai A, Felli IC, Pierattelli R, Uversky VN (2014) The crowd you're in with: Effects of different types of crowding agents on protein aggregation. *Biochim Biophys Acta - Proteins Proteomics* **1844**: 346–357.

290. Furness EL, Ross A, Davis TP, King GC (1998) A hydrophobic interaction site for lysozyme binding to polyethylene glycol and model contact lens polymers. *Biomaterials* **19**: 1361–1369.

291. Wu J, Zhao C, Lin W, Hu R, Wang Q, Chen H, Li L, Chen S, Zheng J (2014) Binding characteristics between polyethylene glycol (PEG) and proteins in aqueous solution. *J Mater Chem B* **2**: 2983–2992.

292. Neale C, Ghanei H, Holyoake J, Bishop RE, Privé GG, Pomès R (2013) Detergent-mediated protein aggregation. *Chem Phys Lipids* **169**: 72–84.

293. Ghosh S, Khatua D, Dey J (2011) Interaction between zwitterionic and anionic surfactants: spontaneous formation of zwitanionic vesicles. *Langmuir* **27**: 5184–5192.

294. Lopez-Moreno R, Fernández-Vivas A, Valverde-Tercedor C, Azuaga Fortes AI, Casares Atienza S, Rodriguez-Navarro AB, Zarivach R, Jimenez-Lopez C (2017) Magnetite nanoparticles biomineralization in the presence of the magnetosome

membrane protein MamC: Effect of protein aggregation and protein structure on magnetite formation. *Cryst Growth Des* **17**: 1620–1629.

295. Ruiz CC (1999) Micelle formation and microenvironmental properties of sodium dodecyl sulfate in aqueous urea solutions. *Colloids Surfaces A Physicochem Eng Asp* **147**: 349–357.

296. Burra G, Thakur AK (2015) Anhydrous trifluoroacetic acid pretreatment converts insoluble polyglutamine peptides to soluble monomers. *Data Br* **5**: 1066–1071.

297. Takeda K, Moriyama Y (1990) Circular dichroism studies on helical structure preferences of amino acid residues of proteins caused by sodium dodecyl sulfate. *J Protein Chem* **9**: 573–582.

298. Sreerama N, Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem* **287**: 252–260.

299. Faham S, Yang D, Bare E, Yohannan S, Whitelegge JP, Bowie JU (2004) Side-chain contributions to membrane protein structure and stability. *J Mol Biol* **335**: 297–305.

300. Blatz PE, Mohler JH, Navangul H V (1972) Anion-induced wavelength regulation of absorption maxima of Schiff bases of retinal. *Biochemistry* **11**: 848–855.

301. Ebrey TG (1971) Energy transfer in rhodopsin, *N*-retinyl-opsin, and rod outer segments. *Proc Natl Acad Sci U S A* **68**: 713–716.

302. Adams RG, Kennedy D, Wulff VJ, Zonana H V. (1958) Rhodopsin bleaching in the presence of hydroxylamine. *Arch Biochem Biophys* **75**: 534–536.

303. Subramaniam S, Marti T, Rosselet SJ, Rothschild KJ, Khorana HG (1991) The reaction of hydroxylamine with bacteriorhodopsin studied with mutants that have altered photocycles: selective reactivity of different photointermediates. *Proc Natl Acad Sci U S A* **88**: 2583–2587.

304. Zadok U, Klare JP, Engelhard M, Sheves M (2005) The hydroxylamine reaction of sensory rhodopsin II: light-induced conformational alterations with C_{13}=C_{14} nonisomerizable pigment. *Biophys J* **89**: 2610–2617.

305. Becher B, Cassim JY (1977) Effects of bleaching and regeneration on the purple membrane structure of *Halobaterium halobium*. *Biophys J* **19**: 285–297.

306. Rafferty CN, Cassim JY, McConnell DG (1977) Circular dichroism, optical rotatory dispersion, and absorption studies on the conformation of bovine rhodopsin *in situ* and solubilized with detergent. *Biophys Struct Mech* **2**: 277–320.

307. Croonen Y, Gelade E, Zegel M Van Der, M. Van der Auweraer, Vandendriessche H, Schryver FC De, Almgren M (1983) Influence of salt, detergent concentration, and temperature on the fluorescence quenching of 1-methylpyrene in sodium dodecyl sulfate with *m*-dicyanobenzene. *J Phys Chem* **87**: 1426–1431.

308. Clint JH (1992) *Surfactant Aggregation*. Springer Netherlands.

309. Schlebach JP, Cao Z, Bowie JU, Park C (2012) Revisiting the folding kinetics of bacteriorhodopsin. *Protein Sci* **21**: 97–106.

310. Silow M, Oliveberg M (1997) Transient aggregates in protein folding are easily

mistaken for folding intermediates. *Proc Natl Acad Sci U S A* **94**: 6084–6086.

311. Cordes EH, Jencks WP (1963) The mechanism of hydrolysis of Schiff bases derived from aliphatic amines. *J Am Chem Soc* **85**: 2843–2848.
312. Lebendiker M, Danieli T (2014) Production of prone-to-aggregate proteins. *FEBS Lett* **588**: 236–246.
313. Munishkina LA, Ahmad A, Fink AL, Uversky VN (2008) Guiding protein aggregation with macromolecular crowding. *Biochemistry* **47**: 8993–9006.
314. Kawamura S, Yokoyama S (1998) Functional characterization of visual and nonvisual pigments of American chameleon (*Anolis carolinensis*). *Vision Res* **38**: 37–44.
315. Pan Y, Brown L, Konermann L (2009) Mapping the structure of an integral membrane protein under semi-denaturing conditions by laser-induced oxidative labeling and mass spectrometry. *J Mol Biol* **394**: 968–981.
316. Okumura H, Murakami M, Kouyama T (2005) Crystal structures of acid blue and alkaline purple forms of bacteriorhodopsin. *J Mol Biol* **351**: 481–495.
317. Booth PJ, Farooq A, Flitsch SL (1996) Retinal binding during folding and assembly of the membrane protein bacteriorhodopsin. *Biochemistry* **35**: 5902–5909.
318. Tastan O, Dutta A, Booth P, Klein-Seetharaman J (2014) Retinal proteins as model systems for membrane protein folding. *Biochim Biophys Acta - Bioenerg* **1837**: 656–663.
319. Sakamoto T, Khorana HG (1995) Structure and function in rhodopsin: the fate of opsin formed upon the decay of light-activated metarhodopsin II *in vitro*. *Proc Natl Acad Sci U S A* **92**: 249–253.
320. Renthal R (2006) An unfolding story of helical transmembrane proteins. *Biochemistry* **45**: 14559–14566.
321. Polet H, Steinhardt J (1968) Binding-induced alterations in ultraviolet absorption of native serum albumin. *Biochemistry* **7**: 1348–1356.
322. Rule GS, Hitchens TK (2006) *Fundamentals of Protein NMR Spectroscopy*. Springer.
323. Williamson MP (2013) Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc* **73**: 1–16.
324. Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease. *Biochemistry* **28**: 8972–8979.
325. Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc* **104**: 4546–4559.
326. Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J Am Chem Soc* **104**: 4559–4570.
327. Fushman D, Weisemann R, Thuring H, Rüterjans H (1994) Backbone dynamics of ribonuclease T1 and its complex with 2′GMP studied by two-dimensional

heteronuclear NMR spectroscopy. *J Biomol NMR* **4**: 61–78.

328. Kneller JM, Lu M, Bracken C (2002) An effective method for the discrimination of motional anisotropy and chemical exchange. *J Am Chem Soc* **124**: 1852–1853.

329. Palmer AG, Massi F (2006) Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem Rev* **106**: 1700–1719.

330. Ishima R (2012) Recent developments in ^{15}N NMR relaxation studies that probe protein backbone dynamics. *Top Curr Chem* **326**: 99–122.

331. Orekhov VY, Pervushin K V., Korzhnev DM, Arseniev AS (1995) Backbone dynamics of (1-71)- and (1-36)bacterioopsin studied by two-dimensional ^{1}H-^{15}N NMR spectroscopy. *J Biomol NMR* **6**: 113–122.

332. Kirkpatrick JP (2007) Solution NMR studies of the helical seven-transmembrane protein pSRII, and NMR structural and dynamical characterisation of the chemokine CTACK (Ph.D. thesis).

333. Le Maire M, Champeil P, Møller J V. (2000) Interaction of membrane proteins and lipids with solubilizing detergents. *Biochim Biophys Acta - Biomembr* **1508**: 86–111.

334. Crowley PB, Kyne C, Monteith WB (2012) Simple and inexpensive incorporation of ^{19}F-tryptophan for protein NMR spectroscopy. *Chem Commun* **48**: 10681–10683.

335. Arntson KE, Pomerantz WCK (2016) Protein-observed fluorine NMR: A bioorthogonal approach for small molecule discovery. *J Med Chem* **59**: 5158–5171.

336. Royant A, Nollert P, Edman K, Neutze R, Landau EM, Pebay-Peyroula E, Navarro J (2001) X-ray structure of sensory rhodopsin II at 2.1-Å resolution. *Proc Nat Acad Sci USA* **98**: 10131–10136.

337. Bostock M (2014) Solution-state NMR spectroscopy method development and functional studies of sensory rhodopsin II (Ph.D. thesis).

338. Schubert M, Kolbe M, Kessler B, Oesterhelt D, Schmieder P (2002) Heteronuclear multidimensional NMR spectroscopy of solubilized membrane proteins: Resonance assignment of native bacteriorhodopsin. *ChemBioChem* **3**: 1019–1023.

339. Klein-Seetharaman J, Reeves PJ, Loewen MC, Getmanova EV, Chung J, Schwalbe H, Wright PE, Khorana HG (2002) Solution NMR spectroscopy of [α-^{15}N]lysine-labeled rhodopsin: The single peak observed in both conventional and TROSY-type HSQC spectra is ascribed to Lys-339 in the carboxyl-terminal peptide sequence. *Proc Natl Acad Sci U S A* **99**: 3452–3457.

340. Klein-Seetharaman J, Yanamala NVK, Javeed F, Reeves PJ, Getmanova EV, Loewen MC, Schwalbe H, Khorana HG (2004) Differential dynamics in the G protein-coupled receptor rhodopsin revealed by solution NMR. *Proc Natl Acad Sci U S A* **101**: 3409–3413.

341. Ishchenko A, Round E, Borshchevskiy V, Grudinin S, Gushchin I, Klare JP, Remeeva A, Polovinkin V, Utrobin P, Balandin T, et al. (2017) New insights on signal propagation by sensory rhodopsin II/transducer complex. *Sci Rep* **7**: 41811.

342. Iwamoto M, Sudo Y, Shimono K, Araiso T, Kamo N (2005) Correlation of the O-intermediate rate with the pKa of Asp-75 in the dark, the counterion of the Schiff base

of *pharaonis* phoborhodopsin (sensory rhodopsin II). *Biophys J* **88**: 1215–1223.

343. Yamaguchi S, Tuzi S, Tanio M, Naito A, Lanyi JK, Needleman R, Saitô H (2000) Irreversible conformational change of bacterio-opsin induced by binding of retinal during its reconstitution to bacteriorhodopsin, as studied by ^{13}C NMR. *J Biochem* **127**: 861–869.

344. Struts A V., Salgado GFJ, Martínez-Mayorga K, Brown MF (2011) Retinal dynamics underlie its switch from inverse agonist to agonist during rhodopsin activation. *Nat Struct Mol Biol* **18**: 392–394.

345. Booth PJ, Riley ML, Flitsch SL, Templer RH, Farooq A, Curran AR, Chadborn N, Wright P (1997) Evidence that bilayer bending rigidity affects membrane protein folding. *Biochemistry* **36**: 197–203.

346. Kahn TW, Engelman DM, Sturtevant JM (1992) Thermodynamic measurements of the contributions of helix-connecting loops and of retinal to the stability of bacteriorhodopsin. *Biochemistry* **31**: 8829–8839.

347. Malmendal A, Underhaug J, Otzen DE, Nielsen NC (2010) Fast mapping of global protein folding states by multivariate NMR: A GPS for proteins. *PLoS One* **5**: e10262.

348. Waudby CA, Ramos A, Cabrita LD, Christodoulou J (2016) Two-dimensional NMR lineshape analysis. *Sci Rep* **6**: 24826.

349. Schanda P, Forge V, Brutscher B (2007) Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy. *Proc Natl Acad Sci U S A* **104**: 11257–11262.

350. Krishnamani V, Hegde BG, Langen R, Lanyi JK (2012) Secondary and tertiary structure of bacteriorhodopsin in the SDS denatured state. *Biochemistry* **51**: 1051–1060.

351. Pan Y, Brown L, Konermann L (2011) Hydrogen/deuterium exchange mass spectrometry and optical spectroscopy as complementary tools for studying the structure and dynamics of a membrane protein. *Int J Mass Spectrom* **302**: 3–11.

352. Khanal A, Pan Y, Brown LS, Konermann L (2012) Pulsed hydrogen/deuterium exchange mass spectrometry for time-resolved membrane protein folding studies. *J Mass Spectrom* **47**: 1620–1626.

353. Hartmann A, Krainer G, Keller S, Schlierf M (2015) Quantification of millisecond protein-folding dynamics in membrane-mimetic environments by single-molecule Förster resonance energy transfer spectroscopy. *Anal Chem* **87**: 11224–11232.

354. Isralewitz B, Izrailev S, Schulten K (1997) Binding pathway of retinal to bacterio-opsin: A prediction by molecular dynamics simulations. *Biophys J* **73**: 2972–2979.

355. Jastrzebska B, Palczewski K, Golczak M (2011) Role of bulk water in hydrolysis of the rhodopsin chromophore. *J Biol Chem* **286**: 18930–18937.

356. Cornilescu G, Hu JS, Bax A (1999) Identification of the hydrogen bonding network in a protein by scalar couplings. *J Am Chem Soc* **121**: 2949–2950.

357. Cordier F, Nisius L, Dingley AJ, Grzesiek S (2008) Direct detection of N-H···O=C hydrogen bonds in biomolecules by NMR spectroscopy. *Nat Protoc* **3**: 235–241.

358. Barfield M (2002) Structural dependencies of interresidue scalar coupling $^{h3}J_{NC'}$ and donor ^{1}H chemical shifts in the hydrogen bonding regions of proteins. *J Am Chem Soc* **124**: 4158–4168.

359. Jaravine VA, Alexandrescu AT, Grzesiek S (2001) Observation of the closing of individual hydrogen bonds during TFE-induced helix formation in a peptide. *Protein Sci* **10**: 943–950.

360. Griesinger C, Sørensen OW, Ernst RR (1986) Correlation of connected transitions by two-dimensional NMR spectroscopy. *J Chem Phys* **85**: 6837–6852.

361. Bax A, Vuister GW, Grzesiek S, Delaglio F, Wang AC, Tschudin R, Zhu G (1994) Measurement of homo- and heteronuclear *J* couplings from quantitative *J* correlation. *Methods Enzymol* **239**: 79–105.

362. Cavanagh J, Fairbrother W, Palmer III A, Rance M, Skelton N (2007) *Protein NMR spectroscopy. Principles and Practice.* Elsevier.

363. Rexroth A, Geppert T, Griesinger C, Schmidt P, Szalma S, Schwalbe H (1995) New principle for the determination of coupling constants that largely suppresses differential relaxation effects. *J Am Chem Soc* **117**: 10389–10390.

364. Pervushin K, Riek R, Wider G, Wuthrich K (1997) Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* **94**: 12366–12371.

365. Tjandra N, Bax A (1997) Measurement of dipolar contributions to $^{1}J_{CH}$ splittings from magnetic-field dependence of *J* modulation in two-dimensional NMR spectra. *J Magn Reson* **124**: 512–515.

366. Werbelow LG, London RE (1995) Dynamic nuclear magnetic resonance frequency shifts for spin 1/2 nuclei coupled to efficiently relaxed spin ≳1/2 nuclei. *J Chem Phys* **102**: 5181–5189.

367. de Alba E, Tjandra N (2006) On the accurate measurement of amide one-bond ^{15}N-^{1}H couplings in proteins: Effects of cross-correlated relaxation, selective pulses and dynamic frequency shifts. *J Magn Reson* **183**: 160–165.

368. de Alba E, Tjandra N (2006) Interference between cross-correlated relaxation and the measurement of scalar and dipolar couplings by Quantitative *J*. *J Biomol NMR* **35**: 1–16.

369. Tjandra N, Grzesiek S, Bax A (1996) Magnetic field dependence of nitrogen-proton *J* splittings in ^{15}N-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. *J Am Chem Soc* **118**: 6264–6272.

370. Yao L, Ying J, Bax A (2009) Improved accuracy of ^{15}N-^{1}H scalar and residual dipolar couplings from gradient-enhanced IPAP-HSQC experiments on protonated proteins. *J Biomol NMR* **43**: 161–170.

371. Burghardt I, Konrat R, Bodenhausen G (1992) Measurement of cross-correlation of fluctuations of dipolar couplings and anisotropic chemical shifts by selective spin locking. *Mol Phys* **75**: 467–486.

372. Tolman JR, Prestegard JH (1996) A quantitative *J*-correlation experiment for the

accurate measurement of one-bond amide ^{15}N-^{1}H couplings in proteins. *J Magn Reson - Ser B* **112**: 245–252.

373. Ottiger M, Delaglio F, Bax A (1998) Measurement of *J* and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson* **131**: 373–378.
374. Shafiq A (2015) Investigating the structural basis of RalA and RalB functional diversity (Ph.D. thesis).
375. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct Funct Genet* **59**: 687–696.
376. Kontaxis G, Clore GM, Bax A (2000) Evaluation of cross-correlation effects and measurement of one-bond couplings in proteins with short transverse relaxation times. *J Magn Reson* **143**: 184–196.
377. Fernández C, Adeishvili K, Wüthrich K (2001) Transverse relaxation-optimized NMR spectroscopy with the outer membrane protein OmpX in dihexanoyl phosphatidylcholine micelles. *Proc Natl Acad Sci U S A* **98**: 2358–2363.
378. Ramsey NF (1953) Electron coupled interactions between nuclear spins in molecules. *Phys Rev* **91**: 303–307.
379. Ma J, Gruschus JM, Tjandra N (2009) ^{15}N-^{1}H scalar coupling perturbation: An additional probe for measuring structural changes due to ligand binding. *J Am Chem Soc* **131**: 9884–9885.
380. Tuttle T, Kraka E, Wu A, Cremer D (2004) Investigation of the NMR spin-spin coupling constants across the hydrogen bonds in ubiquitin: the nature of the hydrogen bond as reflected by the coupling mechanism. *J Am Chem Soc* **126**: 5093–5107.
381. Galiullina LF, Aganova O V., Latfullin IA, Musabirova GS, Aganov A V., Klochkov V V. (2016) NMR study of conformational structure of fluvastatin and its complex with dodecylphosphocholine micelles. *Bionanoscience* **6**: 352–354.
382. Stenson PD, Ball E V., Mort M, Phillips AD, Shiel JA, Thomas NST, Abeysinghe S, Krawczak M, Cooper DN (2003) Human Gene Mutation Database (HGMD®): 2003 update. *Hum Mutat* **21**: 577–581.
383. Tzekov R, Stein L, Kaushal S (2011) Protein misfolding and retinal degeneration. *Cold Spring Harb Perspect Biol* **3**: a007492.
384. Volkmann G, Iwaï H (2010) Protein *trans*-splicing and its use in structural biology: Opportunities and limitations. *Mol Biosyst* **6**: 2110–2121.
385. Saleh L, Perler FB (2006) Protein splicing *in cis* and *in trans*. *Chem Rec* **6**: 183–193.
386. Li Y (2015) Split-inteins and their bioapplications. *Biotechnol Lett* **37**: 2121–2137.
387. Iwai H, Züger S, Jin J, Tam PH (2006) Highly efficient protein *trans*-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. *FEBS Lett* **580**: 1853–1858.
388. Shah NH, Eryilmaz E, Cowburn D, Muir TW (2013) Naturally split inteins assemble through a ‘capture and collapse’ mechanism. *J Am Chem Soc* **135**: 18673–18681.
389. Shi J, Muir TW (2005) Development of a tandem protein *trans*-splicing system based

on native and engineered split inteins. *J Am Chem Soc* **127**: 6198–6206.

390. Shah NH, Vila-Perelló M, Muir TW (2011) Kinetic control of one-pot *trans*-splicing reactions by using a wild-type and designed split intein. *Angew Chemie - Int Ed* **50**: 6511–6515.

391. Xu MQ, Perler FB (1996) The mechanism of protein splicing and its modulation by mutation. *EMBO J* **15**: 5146–5153.

392. Shah NH, Muir TW (2014) Inteins: Nature's gift to protein chemists. *Chem Sci* **5**: 446–461.

393. Frutos S, Goger M, Giovani B, Cowburn D, Muir TW (2010) Branched intermediate formation stimulates peptide bond cleavage in protein splicing. *Nat Chem Biol* **6**: 527–533.

394. Shah NH, Eryilmaz E, Cowburn D, Muir TW (2013) Extein residues play an intimate role in the rate-limiting step of protein *trans*-splicing. *J Am Chem Soc* **135**: 5839–5847.

395. Martin DD, Xu MQ, Evans Jr. TC (2001) Characterization of a naturally occurring *trans*-splicing intein from *Synechocystis* sp. PCC6803. *Biochemistry* **40**: 1393–1402.

396. Cheriyan M, Pedamallu CS, Tori K, Perler F (2013) Faster protein splicing with the *Nostoc punctiforme* DnaE intein using non-native extein residues. *J Biol Chem* **288**: 6202–6211.

397. Southworth MW, Benner J, Perler FB (2000) An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile. *EMBO J* **19**: 5019–5026.

398. Tori K, Dassa B, Johnson MA, Southworth MW, Brace LE, Ishino Y, Pietrokovski S, Perler FB (2010) Splicing of the mycobacteriophage Bethlehem DnaB intein: Identification of a new mechanistic class of inteins that contain an obligate block F nucleophile. *J Biol Chem* **285**: 2515–2526.

399. Giaccone Z, Reitter J, Steeves A, Kulik H, Mills K (2015) Peptide bond cleavage through asparagine cyclization. *FASEB J* **29**: 722.9.

400. Muona M, Aranko AS, Raulinaitis V, Iwaï H (2010) Segmental isotopic labeling of multi-domain and fusion proteins by protein *trans*-splicing *in vivo* and *in vitro*. *Nat Protoc* **5**: 574–587.

401. Zettler J, Schütz V, Mootz HD (2009) The naturally split *Npu* DnaE intein exhibits an extraordinarily high rate in the protein *trans*-splicing reaction. *FEBS Lett* **583**: 909–914.

402. Züger S, Iwai H (2005) Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. *Nat Biotechnol* **23**: 736–740.

403. Giriat I, Muir TW (2003) Protein semi-synthesis in living cells. *J Am Chem Soc* **125**: 7180–7181.

404. Kwon Y, Coleman MA, Camarero JA (2006) Selective immobilization of proteins onto solid supports through split-intein-mediated protein *trans*-splicing. *Angew Chemie - Int Ed* **45**: 1726–1729.

405. Scott CP, Abel-Santos E, Wall M, Wahnon DC, Benkovic SJ (1999) Production of cyclic peptides and proteins *in vivo*. *Proc Natl Acad Sci U S A* **96**: 13638–13643.

406. Mills K V., Manning JS, Garcia AM, Wuerdeman LA (2004) Protein splicing of a *Pyrococcus abyssi* intein with a C-terminal glutamine. *J Biol Chem* **279**: 20685–20691.

407. Brenzel S, Kurpiers T, Mootz HD (2006) Engineering artificially split inteins for applications in protein chemistry: Biochemical characterization of the split *Ssp* DnaB intein and comparison to the split *Sce* VMA intein. *Biochemistry* **45**: 1571–1578.

408. Al-Ali H, Ragan TJ, Gao X, Harris TK (2007) Reconstitution of modular PDK1 functions on *trans*-splicing of the regulatory PH and catalytic kinase domains. *Bioconjug Chem* **18**: 1294–1302.

409. Muona M, Aranko AS, Iwai H (2008) Segmental isotopic labelling of a multidomain protein by protein ligation by protein *trans*-splicing. *ChemBioChem* **9**: 2958–2961.

410. Zhou P, Lugovskoy AA, Wagner G (2001) A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins. *J Biomol NMR* **20**: 11–14.

411. Nallamsetty S, Waugh DS (2006) Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expr Purif* **45**: 175–182.

412. Etzkorn M, Seidel K, Li L, Martell S, Geyer M, Engelhard M, Baldus M (2010) Complex formation and light activation in membrane-embedded sensory rhodopsin II as seen by solid-state NMR spectroscopy. *Structure* **18**: 293–300.

413. Yang CS, Sineshchekov O, Spudich EN, Spudich JL (2004) The cytoplasmic membrane-proximal domain of the HtrII transducer interacts with the E-F loop of photoactivated *Natronomonas pharaonis* sensory rhodopsin II. *J Biol Chem* **279**: 42970–42976.

414. Shimono K, Hayashi T, Ikeura Y, Sudo Y, Iwamoto M, Kamo N (2003) Importance of the broad regional interaction for spectral tuning in *Natronobacterium pharaonis* phoborhodopsin (sensory rhodopsin II). *J Biol Chem* **278**: 23882–23889.

415. Lüneberg J, Widmann M, Dathe M, Marti T (1998) Secondary structure of bacteriorhodopsin fragments: External sequence constraints specify the conformation of transmembrane helices. *J Biol Chem* **273**: 28822–28830.

416. Hsu MF, Fu HY, Cai CJ, Yi HP, Yang CS, Wang AHJ (2015) Structural and functional studies of a newly grouped *Haloquadratum walsbyi* bacteriorhodopsin reveal the acid-resistant light-driven proton pumping activity. *J Biol Chem* **290**: 29567–29577.

417. Pebay-Peyroula E, Royant A, Landau EM, Navarro J (2002) Structural basis for sensory rhodopsin function. *Biochim Biophys Acta - Biomembr* **1565**: 196–205.

418. Lythgoe JN (1972) List of Vertebrate Visual Pigments. In Dartnall HJA (ed.), *Photochemistry of Vision* pp 604–624. Springer Berlin Heidelberg, Berlin, Heidelberg.

419. Miroux B, Walker JE (1996) Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* **260**: 289–298.

420. Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, Högbom M, van Wijk KJ, Slotboom DJ, Persson JO, et al. (2008) Tuning *Escherichia coli* for membrane protein overexpression. *Proc Natl Acad Sci U S A* **105**: 14371–14376.

421. Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T (1998) Chaperone coexpression plasmids: Differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl Environ Microbiol* **64**: 1694–1699.

422. Aranko AS, Züger S, Buchinger E, Iwaï H (2009) *In vivo* and *in vitro* protein ligation by naturally occurring and engineered split DnaE inteins. *PLoS One* **4**: e5185.

423. Gunasekaran K, Eyles SJ, Hagler AT, Gierasch LM (2001) Keeping it in the family: folding studies of related proteins. *Curr Opin Struct Biol* **11**: 83–93.

424. Zarrine-Afsar A, Larson SM, Davidson AR (2005) The family feud: do proteins with similar structures fold via the same pathway? *Curr Opin Struct Biol* **15**: 42–49.

425. Shen L, Chen C, Zheng H, Jin L (2013) The evolutionary relationship between microbial rhodopsins and metazoan rhodopsins. *Sci World J* **2013**: 435651.

426. Maxwell KL, Wildes D, Zarrine-Afsar A, De Los Rios MA, Brown AG, Friel CT, Hedberg L, Horng J-C, Bona D, Miller EJ, et al. (2005) Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci* **14**: 602–616.

427. Seddon AM, Curnow P, Booth PJ (2004) Membrane proteins, lipids and detergents: Not just a soap opera. *Biochim Biophys Acta - Biomembr* **1666**: 105–117.

428. Tastan O, Klein-Seetharaman J, Meirovitch H (2009) The effect of loops on the structural organization of α-helical membrane proteins. *Biophys J* **96**: 2299–2312.

429. Baumann A, Kerruth S, Fitter J, Büldt G, Heberle J, Schlesinger R, Ataka K (2016) *In-situ* observation of membrane protein folding during cell-free expression. *PLoS One* **11**: e0151051.

430. Bischoff L, Wickles S, Berninghausen O, Van Der Sluis EO, Beckmann R (2014) Visualization of a polytopic membrane protein during SecY-mediated membrane insertion. *Nat Commun* **5**: 4103.

431. Sambrook J, Russell DW (2006) Preparation and transformation of competent *E. coli* using calcium chloride. *Cold Spring Harb Protoc* **2006**: pdb.prot3932.

432. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680–685.

433. Hochuli M, Szyperski T, Wüthrich K (2000) Deuterium isotope effects on the central carbon metabolism of *Escherichia coli* cells grown on a D_2O-containing minimal medium. *J Biomol NMR* **17**: 33–42.

434. Scharf B, Hess B, Engelhard M (1992) Chromophore of sensory rhodopsin II from Halobacterium halobium. *Biochemistry* **31**: 12486–12492.

435. Provencher SW, Glockner J (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20**: 33–37.

436. Sreerama N, Woody RW (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal Biochem* **209**: 32–44.

437. Manavalan P, Johnson WC (1987) Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal Biochem* **167**: 76–

85.

438. Mitchell J (2018) Structural stability, dynamics and unfolding of 7-transmembrane helical receptors (Ph.D. thesis).

439. Elzhov T V, Mullen KM, Spiess A-N, Bolker B (2016) *Package 'minpack.lm': R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds*.

440. R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*.